Quantum Mechanics

Quantum Mechanics

Fourth edition

Alastair I. M. Rae

Department of Physics
University of Birmingham
UK

IoP

Institute of Physics Publishing
Bristol and Philadelphia

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 0839 7

Library of Congress Cataloging-in-Publication Data are available

First edition 1980
Second edition 1986
Reprinted 1987
Reprinted with corrections 1990
Reprinted 1991
Third edition 1992
Reprinted 1993
Reprinted with corrections 1996
Reprinted 1998, 2001
Fourth edition 2002

Commissioning Editor: James Revill
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Frédérique Swist

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 1035, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset in the UK by Text 2 Text, Torquay, Devon
Printed in the UK by MPG Books Ltd, Bodmin, Cornwall

To Angus and Gavin

Contents

Preface to Fourth Edition

When I told a friend that I was working on a new edition, he asked me what had changed in quantum physics during the last ten years. In one sense very little: quantum mechanics is a very well established theory and the basic ideas and concepts are little changed from what they were ten, twenty or more years ago. However, new applications have been developed and some of these have revealed aspects of the subject that were previously unknown or largely ignored. Much of this development has been in the field of information processing, where quantum effects have come to the fore. In particular, quantum techniques appear to have great potential in the field of cryptography, both in the coding and possible de-coding of messages, and I have included a chapter aimed at introducing this topic.

I have also added a short chapter on relativistic quantum mechanics and introductory quantum field theory. This is a little more advanced than many of the other topics treated, but I hope it will be accessible to the interested reader. It aims to open the door to the understanding of a number of points that were previously stated without justification.

Once again, I have largely re-written the last chapter on the conceptual foundations of the subject. The twenty years since the publication of the first edition do not seem to have brought scientists and philosophers significantly closer to a consensus on these problems. However, many issues have been considerably clarified and the strengths and weaknesses of some of the explanations are more apparent. My own understanding continues to grow, not least because of what I have learned from formal and informal discussions at the annual UK Conferences on the Foundations of Physics.

Other changes include a more detailed treatment of tunnelling in chapter 2, a more gentle transition from the Born postulate to quantum measurement theory in chapter 4, the introduction of Dirac notation in chapter 6 and a discussion of the Bose–Einstein condensate in chapter 10.

I am grateful to a number of people who have helped me with this edition. Glenn Cox shared his expertise on relativistic quantum mechanics when he read a draft of chapter 11; Harvey Brown corrected my understanding of the de Broglie–Bohm hidden variable theory discussed in the first part of chapter 13; Demetris Charalambous read a late draft of the whole book and suggested several

improvements and corrections. Of course, I bear full responsibility for the final version and any remaining errors.

Modern technology means that the publishers are able to support the book at the web site http://bookmarkphysics.iop.org/bookpge.htm/book=1107p. This is where you will find references to the wider literature, colour illustrations, links to other relevant web sites, etc. If any mistakes are identified, corrections will also be listed there. Readers are also invited to contribute suggestions on what would be useful content. The most convenient form of communication is by e-mail to 0750308397@bookmarkphysics.iop.org.

Finally I should like to pay tribute to Ann for encouraging me to return to writing after some time. Her support has been invaluable.

Alastair I. M. Rae
2002

Preface to Third Edition

In preparing this edition, I have again gone right through the text identifying points where I thought the clarity could be improved. As a result, numerous minor changes have been made. More major alterations include a discussion of the impressive modern experiments that demonstrate neutron diffraction by macroscopic sized slits in chapter 1, a revised treatment of Clebsch–Gordan coefficients in chapter 6 and a fuller discussion of spontaneous emission in chapter 8. I have also largely rewritten the last chapter on the conceptual problems of quantum mechanics in the light of recent developments in the field as well as of improvements in my understanding of the issues involved and changes in my own viewpoint. This chapter also includes an introduction to the de Broglie–Bohm hidden variable theory and I am grateful to Chris Dewdney for a critical reading of this section.

Alastair I. M. Rae
1992

Preface to Second Edition

I have not introduced any major changes to the structure or content of the book, but I have concentrated on clarifying and extending the discussion at a number of points. Thus the discussion of the application of the uncertainty principle to the Heisenberg microscope has been revised in chapter 1 and is referred to again in chapter 4 as one of the examples of the application of the generalized uncertainty principle; I have rewritten much of the section on spin–orbit coupling and the Zeeman effect and I have tried to improve the introduction to degenerate perturbation theory which many students seem to find difficult. The last chapter has been brought up to date in the light of recent experimental and theoretical work on the conceptual basis of the subject and, in response to a number of requests from students, I have provided hints to the solution of the problems at the ends of the chapters.

I should like to thank everyone who drew my attention to errors or suggested improvements, I believe nearly every one of these suggestions has been incorporated in one way or another into this new edition.

Alastair I. M. Rae
1985

Preface to First Edition

Over the years the emphasis of undergraduate physics courses has moved away from the study of classical macroscopic phenomena towards the discussion of the microscopic properties of atomic and subatomic systems. As a result, students now have to study quantum mechanics at an earlier stage in their course without the benefit of a detailed knowledge of much of classical physics and, in particular, with little or no acquaintance with the formal aspects of classical mechanics. This book has been written with the needs of such students in mind. It is based on a course of about thirty lectures given to physics students at the University of Birmingham towards the beginning of their second year—although, perhaps inevitably, the coverage of the book is a little greater than I was able to achieve in the lecture course. I have tried to develop the subject in a reasonably rigorous way, covering the topics needed for further study in atomic, nuclear, and solid state physics, but relying only on the physical and mathematical concepts usually taught in the first year of an undergraduate course. On the other hand, by the end of their first undergraduate year most students have heard about the basic ideas of atomic physics, including the experimental evidence pointing to the need for a quantum theory, so I have confined my treatment of these topics to a brief introductory chapter.

While discussing these aspects of quantum mechanics required for further study, I have laid considerable emphasis on the understanding of the basic ideas and concepts behind the subject, culminating in the last chapter which contains an introduction to quantum measurement theory. Recent research, particularly the theoretical and experimental work inspired by Bell's theorem, has greatly clarified many of the conceptual problems in this area. However, most of the existing literature is at a research level and concentrates more on a rigorous presentation of results to other workers in the field than on making them accessible to a wider audience. I have found that many physics undergraduates are particularly interested in this aspect of the subject and there is therefore a need for a treatment suitable for this level. The last chapter of this book is an attempt to meet this need.

I should like to acknowledge the help I have received from my friends and colleagues while writing this book. I am particularly grateful to Robert Whitworth, who read an early draft of the complete book, and to Goronwy Jones and George Morrison, who read parts of it. They all offered many valuable and

penetrating criticisms, most of which have been incorporated in this final version. I should also like to thank Ann Aylott who typed the manuscript and was always patient and helpful throughout many changes and revisions, as well as Martin Dove who assisted with the proofreading. Naturally, none of this help in any way lessens my responsibility for whatever errors and omissions remain.

Alastair I. M. Rae
1980

Chapter 1

Introduction

Quantum mechanics was developed as a response to the inability of the classical theories of mechanics and electromagnetism to provide a satisfactory explanation of some of the properties of electromagnetic radiation and of atomic structure. As a result, a theory has emerged whose basic principles can be used to explain not only the structure and properties of atoms, molecules and solids, but also those of nuclei and of 'elementary' particles such as the proton and neutron. Although there are still many features of the physics of such systems that are not fully understood, there are presently no indications that the fundamental ideas of quantum mechanics are incorrect. In order to achieve this success, quantum mechanics has been built on a foundation that contains a number of concepts that are fundamentally different from those of classical physics and which have radically altered our view of the way the natural universe operates. This book aims to elucidate and discuss the conceptual basis of the subject as well as explaining its success in describing the behaviour of atomic and subatomic systems.

Quantum mechanics is often thought to be a difficult subject, not only in its conceptual foundation, but also in the complexity of its mathematics. However, although a rather abstract formulation is required for a proper treatment of the subject, much of the apparent complication arises in the course of the solution of essentially simple mathematical equations applied to particular physical situations. We shall discuss a number of such applications in this book, because it is important to appreciate the success of quantum mechanics in explaining the results of real physical measurements. However, the reader should try not to allow the ensuing algebraic complication to hide the essential simplicity of the basic ideas.

In this first chapter we shall discuss some of the key experiments that illustrate the failure of classical physics. However, although the experiments described were performed in the first quarter of this century and played an important role in the development of the subject, we shall not be giving a historically based account. Neither will our account be a complete description of the early experimental work. For example, we shall not describe the experiments

on the properties of thermal radiation and the heat capacity of solids that provided early indications of the need for the quantization of the energy of electromagnetic radiation and of mechanical systems. The topics to be discussed have been chosen as those that point most clearly towards the basic ideas needed in the further development of the subject. As so often happens in physics, the way in which the theory actually developed was by a process of trial and error, often relying on flashes of inspiration, rather than the possibly more logical approach suggested by hindsight.

1.1 The photoelectric effect

When light strikes a clean metal surface in a vacuum, it causes electrons to be emitted with a range of energies. For light of a given frequency ν the maximum electron energy E_x is found to be equal to the difference between two terms. One of these is proportional to the frequency of the incident light with a constant of proportionality h that is the same whatever the metal used, while the other is independent of frequency but varies from metal to metal. Neither term depends on the intensity of the incident light, which affects only the rate of electron emission. Thus

$$E_x = h\nu - \phi \tag{1.1}$$

It is impossible to explain this result on the basis of the classical theory of light as an electromagnetic wave. This is because the energy contained in such a wave would arrive at the metal at a uniform rate and there is no apparent reason why this energy should be divided up in such a way that the maximum electron energy is proportional to the frequency and independent of the intensity of the light. This point is emphasized by the dependence of the rate of electron emission on the light intensity. Although the average emission rate is proportional to the intensity, individual electrons are emitted at random. It follows that electrons are sometimes emitted well before sufficient electromagnetic energy should have arrived at the metal, and this point has been confirmed by experiments performed using very weak light.

Such considerations led Einstein to postulate that the classical electromagnetic theory does not provide a complete explanation of the properties of light, and that we must also assume that the energy in an electromagnetic wave is 'quantized' in the form of small packets, known as *photons*, each of which carries an amount of energy equal to $h\nu$. Given this postulate, we can see that when light is incident on a metal, the maximum energy an electron can gain is that carried by one of the photons. Part of this energy is used to overcome the binding energy of the electron to the metal—so accounting for the quantity ϕ in (1.1), which is known as the *work function*. The rest is converted into the kinetic energy of the freed electron, in agreement with the experimental results summarized in equation (1.1). The photon postulate also explains the emission of photoelectrons at random times. Thus, although the average rate of photon arrival is proportional to

the light intensity, individual photons arrive at random and, as each carries with it a quantum of energy, there will be occasions when an electron is emitted well before this would be classically expected.

The constant h connecting the energy of a photon with the frequency of the electromagnetic wave is known as *Planck's constant*, because it was originally postulated by Max Planck in order to explain some of the properties of thermal radiation. It is a fundamental constant of nature that frequently occurs in the equations of quantum mechanics. We shall find it convenient to change this notation slightly and define another constant $\hbar$ as being equal to h divided by 2π. Moreover, when referring to waves, we shall normally use the angular frequency $\omega(= 2\pi\nu)$, in preference to the frequency ν. Using this notation, the photon energy E can be expressed as

$$E = \hbar\omega \tag{1.2}$$

Throughout this book we shall write our equations in terms of $\hbar$ and avoid ever again referring to h. We note that $\hbar$ has the dimensions of energy×time and its currently best accepted value is $1.054\,571\,596 \times 10^{-34}$ J s.

1.2 The Compton effect

The existence of photons is also demonstrated by experiments involving the scattering of x-rays by electrons, which were first carried out by A. H. Compton. To understand these we must make the further postulate that a photon, as well as carrying a quantum of energy, also has a definite momentum and can therefore be treated in many ways just like a classical particle. An expression for the photon momentum is suggested by the classical theory of radiation pressure: it is known that if energy is transported by an electromagnetic wave at a rate W per unit area per second, then the wave exerts a pressure of magnitude W/c (where c is the velocity of light), whose direction is parallel to that of the wavevector $\mathbf{k}$ of the wave; if we now treat the wave as composed of photons of energy $\hbar\omega$ it follows that the photon momentum $\mathbf{p}$ should have a magnitude $\hbar\omega/c = \hbar k$ and that its direction should be parallel to $\mathbf{k}$. Thus

$$\mathbf{p} = \hbar\mathbf{k} \tag{1.3}$$

We now consider a collision between such a photon and an electron of mass m that is initially at rest. After the collision we assume that the frequency and wavevector of the photon are changed to ω' and $\mathbf{k}'$ and that the electron moves off with momentum $\mathbf{p}_e$ as shown in figure 1.1. From the conservation of energy and momentum, we have

$$\hbar\omega - \hbar\omega' = p_e^2/2m \tag{1.4}$$

$$\hbar\mathbf{k} - \hbar\mathbf{k}' = \mathbf{p}_e \tag{1.5}$$

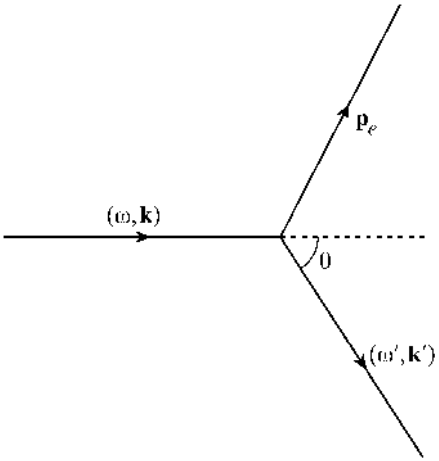


Figure 1.1. In Compton scattering an x-ray photon of angular frequency ω and wavevector $\mathbf{k}$ collides with an electron initially at rest. After the collision the photon frequency and wavevector are changed to ω' and $\mathbf{k}'$ respectively and the electron recoils with momentum $\mathbf{p}_e$.

Squaring (1.5) and substituting into (1.4) we get

$$\begin{aligned}\hbar(\omega - \omega') &= \frac{\hbar^2}{2m}(\mathbf{k} - \mathbf{k}')^2 = \frac{\hbar^2}{2m}[k^2 + k'^2 - 2kk'\cos\theta] \\ &= \frac{\hbar^2}{2m}[(k - k')^2 + 2kk'(1 - \cos\theta)] \end{aligned} \tag{1.6}$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$ (cf. figure 1.1). Now the change in the magnitude of the wavevector $(k - k')$ always turns out to be very much smaller than either k or k' so we can neglect the first term in square brackets on the right-hand side of (1.6). Remembering that $\omega = ck$ and $\omega' = ck'$ we then get

$$\frac{1}{\omega'} - \frac{1}{\omega} = \frac{\hbar}{mc^2}(1 - \cos\theta)$$

that is

$$\lambda' - \lambda = \frac{2\pi\hbar}{mc}(1 - \cos\theta) \tag{1.7}$$

where λ and λ' are the x-ray wavelengths before and after the collision, respectively. It turns out that if we allow for relativistic effects when carrying out this calculation, we obtain the same result as (1.7) without having to make any approximations.

Experimental studies of the scattering of x-rays by electrons in solids produce results in good general agreement with these predictions. In particular,

if the intensity of the radiation scattered through a given angle is measured as a function of the wavelength of the scattered x-rays, a peak is observed whose maximum lies just at the point predicted by (1.7). In fact such a peak has a finite, though small, width implying that some of the photons have been scattered in a manner slightly different from that described above. This can be explained by taking into account the fact that the electrons in a solid are not necessarily at rest, but generally have a finite momentum before the collision. Compton scattering can therefore be used as a tool to measure the electron momentum, and we shall discuss this in more detail in chapter 4.

Both the photoelectric effect and the Compton effect are connected with the interactions between electromagnetic radiation and electrons, and both provide conclusive evidence for the photon nature of electromagnetic waves. However, we might ask why there are two effects and why the x-ray photon is scattered by the electron with a change of wavelength, while the optical photon transfers all its energy to the photoelectron. The principal reason is that in the x-ray case the photon energy is much larger than the binding energy between the electron and the solid; the electron is therefore knocked cleanly out of the solid in the collision and we can treat the problem by considering energy and momentum conservation. In the photoelectric effect, on the other hand, the photon energy is only a little larger than the binding energy and, although the details of this process are rather complex, it turns out that the momentum is shared between the electron and the atoms in the metal and that the whole of the photon energy can be used to free the electron and give it kinetic energy. However, none of these detailed considerations affects the conclusion that in both cases the incident electromagnetic radiation exhibits properties consistent with it being composed of photons whose energy and momentum are given by the expressions (1.2) and (1.3).

1.3 Line spectra and atomic structure

When an electric discharge is passed through a gas, light is emitted which, when examined spectroscopically, is typically found to consist of a series of lines, each of which has a sharply defined frequency. A particularly simple example of such a line spectrum is that of hydrogen, in which case the observed frequencies are given by the formula

$$\omega_{mn} = 2\pi R_0 c \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \tag{1.8}$$

where n and m are integers, c is the speed of light and R_0 is a constant known as the *Rydberg constant* (after J. R. Rydberg who first showed that the experimental results fitted this formula) whose currently accepted value is $1.097\,373\,157 \times 10^7\ \mathrm{m}^{-1}$.

Following our earlier discussion, we can assume that the light emitted from the atom consists of photons whose energies are $\hbar\omega_{mn}$. It follows from this and

the conservation of energy that the energy of the atom emitting the photon must have been changed by the same amount. The obvious conclusion to draw is that the energy of the hydrogen atom is itself quantized, meaning that it can adopt only one of the values E_n where

$$E_n = -\frac{2\pi R_0 \hbar c}{n^2} \tag{1.9}$$

the negative sign corresponding to the negative binding energy of the electron in the atom. Similar constraints govern the values of the energies of atoms other than hydrogen, although these cannot usually be expressed in such a simple form. We refer to allowed energies such as E_n as *energy levels*. Further confirmation of the existence of energy levels is obtained from the ionization energies and absorption spectra of atoms, which both display features consistent with the energy of an atom being quantized in this way. It will be one of the main aims of this book to develop a theory of quantum mechanics that will successfully explain the existence of energy levels and provide a theoretical procedure for calculating their values.

One feature of the structure of atoms that can be at least partly explained on the basis of energy quantization is the simple fact that atoms exist at all! According to classical electromagnetic theory, an accelerated charge always loses energy in the form of radiation, so a negative electron in motion about a positive nucleus should radiate, lose energy, and quickly coalesce with the nucleus. The fact that the radiation is quantized should not affect this argument, but if the energy of the atom is quantized, there will be a minimum energy level (that with $n = 1$ in the case of hydrogen) below which the atom cannot go, and in which it will remain indefinitely. Quantization also explains why all atoms of the same species behave in the same way. As we shall see later, all hydrogen atoms in the lowest energy state have the same properties. This is in contrast to a classical system, such as a planet orbiting a star, where an infinite number of possible orbits with very different properties can exist for a given value of the energy of the system.

1.4 de Broglie waves

Following on from the fact that the photons associated with electromagnetic waves behave like particles, L. de Broglie suggested that particles such as electrons might also have wave properties. He further proposed that the frequencies and wavevectors of these 'matter waves' would be related to the energy and momentum of the associated particle in the same way as in the photon case. That is

$$\left.\begin{aligned} E &= \hbar\omega \\ \mathbf{p} &= \hbar\mathbf{k} \end{aligned}\right\} \tag{1.10}$$

In the case of matter waves, equations (1.10) are referred to as the de Broglie relations. We shall develop this idea in subsequent chapters, where we shall find that it leads to a complete description of the structure and properties of atoms, including the quantized atomic energy levels. In the meantime we shall describe an experiment that provides direct confirmation of the existence of matter waves.

The property possessed by a wave that distinguishes it from any other physical phenomenon is its ability to form interference and diffraction patterns: when different parts of a wave are recombined after travelling different distances, they reinforce each other or cancel out depending on whether the two path lengths differ by an even or an odd number of wavelengths. Such phenomena are readily demonstrated in the laboratory by passing light through a diffraction grating for example. However, if the wavelength of the waves associated with even very low energy electrons (say around 1 eV) is calculated using the de Broglie relations (1.10) a value of around 10^{-9} m is obtained, which is much smaller than that of visible light and much too small to form a detectable diffraction pattern when passed through a conventional grating. However, the atoms in a crystal are arranged in periodic arrays, so a crystal can act as a three-dimensional diffraction grating with a very small spacing. This is demonstrated in x-ray diffraction, and the first direct confirmation of de Broglie's hypothesis was an experiment performed by C. Davisson and L. H. Germer that showed electrons being diffracted by crystals in a similar manner.

Nowadays the wave properties of electron beams are commonly observed experimentally and electron microscopes, for example, are often used to display the diffraction patterns of the objects under observation. Moreover, not only electrons behave in this way; neutrons of the appropriate energy can also be diffracted by crystals, this technique being commonly used to investigate structural and other properties of solids. In recent years, neutron beams have been produced with such low energy that their de Broglie wavelength is as large as 2.0 nm. When these are passed through a double slit whose separation is of the order of 0.1 mm, the resulting diffraction maxima are separated by about 10^{-3} degrees, which corresponds to about 0.1 mm at a distance of 5 m beyond the slits, where the detailed diffraction pattern can be resolved. Figure 1.2 gives the details of such an experiment and the results obtained; we see that the number of neutrons recorded at different angles is in excellent agreement with the intensity of the diffraction pattern, calculated on the assumption that the neutron beam can be represented by a de Broglie wave.

1.5 Wave–particle duality

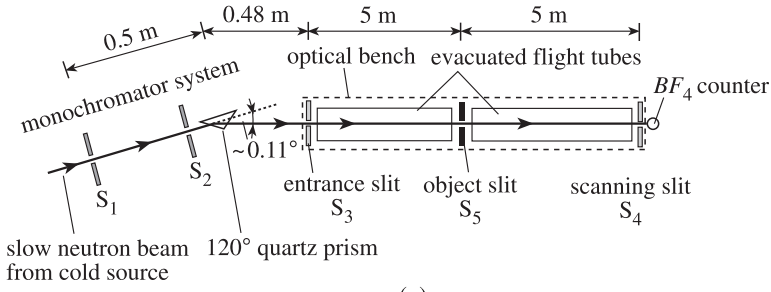
Although we have just described the experimental evidence for the wave nature of electrons and similar bodies, it must not be thought that this description is complete or that these are any-the-less particles. Although in a diffraction experiment wave properties are manifested during the diffraction process and the

intensity of the wave determines the average number of particles scattered through various angles, when the diffracted electrons are detected they are always found to behave like point particles with the expected mass and charge and having a particular energy and momentum. Conversely, although we need to postulate photons in order to explain the photoelectric and Compton effects, phenomena such as the diffraction of light by a grating or of x-rays by a crystal can be explained only if electromagnetic radiation has wave properties.

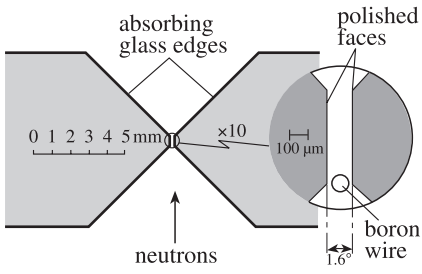
Quantum mechanics predicts that both the wave and the particle models apply to all objects whatever their size. However, in many circumstances it is perfectly clear which model should be used in a particular physical situation. Thus, electrons with a kinetic energy of about 100 eV (1.6×10^{-17} J) have a de Broglie wavelength of about 10^{-10} m and are therefore diffracted by crystals according to the wave model. However, if their energy is very much higher (say 100 MeV) the wavelength is then so short (about 10^{-14} m) that diffraction effects are not normally observed and such electrons nearly always behave like classical particles. A small grain of sand of mass about 10^{-6} g moving at a speed of 10^{-3} m s^{-1} has a de Broglie wavelength of the order of 10^{-21} m and its wave properties are quite undetectable; clearly this is even more true for heavier or faster moving objects. There is considerable interest in attempting to detect wave properties of more and more massive objects. To date, the heaviest body for which diffraction of de Broglie waves has been directly observed is the Buckminster fullerene molecule C_{60} whose mass is nearly 1000 times that of a neutron. These particles were passed through a grating and the resulting diffraction pattern was observed in an experiment performed in 2000 by the same group as is featured in figure 1.2.

Some experiments cannot be understood unless the wave and particle are both used. If we examine the neutron diffraction experiment illustrated in figure 1.2, we see how it illustrates this. The neutron beam behaves like a wave when it is passing through the slits and forming an interference pattern, but when the neutrons are detected, they behave like a set of individual particles with the usual mass, zero electric charge etc. We never detect half a neutron! Moreover, the typical neutron beams used in such experiments are so weak that no more than one neutron is in the apparatus at any one time and we therefore cannot explain the interference pattern on the basis of any model involving interactions between different neutrons.

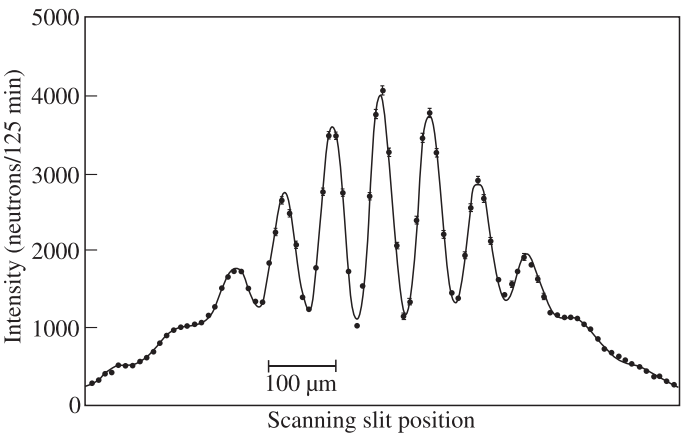
Suppose we now change this experiment by placing detectors behind each slit instead of a large distance away; these will detect individual neutrons passing through one or other of the slits—but never both at once—and the obvious conclusion is that the same thing happened in the interference experiment. But we have just seen that the interference pattern is formed by a wave passing through both slits, and this can be confirmed by arranging a system of shutters so that only one or other of the two slits, but never both, are open at any one time, in which case it is impossible to form an interference pattern. Both slits are necessary to form the interference pattern, so if the neutrons always pass through one slit or



(*a*)



(*b*)



(*c*)

Figure 1.2. In recent years, it has been possible to produce neutron beams with de Broglie wavelengths around 2 nm which can be detectably diffracted by double slits of separation about 0.1 mm. A typical experimental arrangement is shown in (*a*) and the slit arrangement is illustrated in (*b*). The number of neutrons recorded along a line perpendicular to the diffracted beam 5 m beyond the slits is shown in (*c*), along with the intensity calculated from diffraction theory, assuming a wave model for the neutron beam. The agreement is clearly excellent. (Reproduced by permission from A. Zeilinger, R. Gähler, C. G. Schull, W. Treimer and W. Mampe, *Reviews of Modern Physics* **60** 1067–73 (1988).)

the other then the behaviour of a given neutron must somehow be affected by the slit it did not pass through!

An alternative view, which is now the orthodox interpretation of quantum mechanics, is to say that the model we use to describe quantum phenomena is not just a property of the quantum objects (the neutrons in this case) but also depends on the arrangement of the whole apparatus. Thus, if we perform a diffraction experiment, the neutrons are waves when they pass through the slits, but are particles when they are detected. But if the experimental apparatus includes detectors right behind the slits, the neutrons behave like particles at this point. This dual description is possible because no interference pattern is created in the latter case. Moreover, it turns out that this happens no matter how subtle an experiment we design to detect which slit the neutron passes through: if it is successful, the phase relation between the waves passing through the slits is destroyed and the interference pattern disappears. We can therefore look on the particle and wave models as *complementary* rather than contradictory properties. Which one is manifest in a particular experimental situation depends on the arrangement of the whole apparatus, including the slits and the detectors; we should not assume that, just because we detect particles when we place detectors behind the slits, the neutrons still have these properties when we do not.

It should be noted that, although we have just discussed neutron diffraction, the argument would have been largely unchanged if we had considered light waves and photons or any other particles with their associated waves. In fact the idea of complementarity is even more general than this and we shall find many cases in our discussion of quantum mechanics where the measurement of one property of a physical system renders another unobservable; an example of this will be described in the next paragraph when we discuss the limitations on the simultaneous measurement of the position and momentum of a particle. Many of the apparent paradoxes and contradictions that arise can be resolved by concentrating on those aspects of a physical system that can be directly observed and refraining from drawing conclusions about properties that cannot. However, there are still significant conceptual problems in this area which remain the subject of active research, and we shall discuss these in some detail in chapter 13.

The uncertainty principle

In this section we consider the limits that wave–particle duality places on the simultaneous measurement of the position and momentum of a particle. Suppose we try to measure the position of a particle by illuminating it with radiation of wavelength λ and using a microscope of angular aperture α, as shown in figure 1.3. The fact that the radiation has wave properties means that the size of the image observed in the microscope will be governed by the resolving power of the microscope. The position of the electron is therefore uncertain by an amount

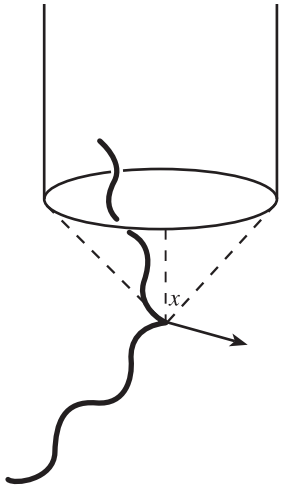


Figure 1.3. A measurement of the position of a particle by a microscope causes a corresponding uncertainty in the particle momentum as it recoils after interaction with the illuminating radiation.

Δx which is given by standard optical theory as

$$\Delta x \simeq \frac{\lambda}{\sin\alpha} \tag{1.11}$$

However, the fact that the radiation is composed of photons means that each time the particle is struck by a photon it recoils, as in Compton scattering. The momentum of the recoil could of course be calculated if we knew the initial and final momenta of the photon, but as we do not know through which points on the lens the photons entered the microscope, the x component of the particle momentum is subject to an error Δp_x where

$$\begin{aligned} \Delta p_x &\simeq p \sin\alpha \\ &= 2\pi\hbar \sin\alpha/\lambda \end{aligned} \tag{1.12}$$

Combining (1.11) and (1.12) we get

$$\Delta x \Delta p_x \simeq 2\pi\hbar \tag{1.13}$$

It follows that if we try to improve the accuracy of the position measurement by using radiation with a smaller wavelength, we shall increase the error on the momentum measurement and vice versa. This is just one example of an experiment designed to measure the position and momentum of a particle, but it turns out that any other experiment with this aim is subject to constraints similar to (1.13). We shall see in chapter 4 that the fundamental principles of

quantum mechanics ensure that in every case the uncertainties in the position and momentum components are related by

$$\Delta x \, \Delta p_x \geqslant \tfrac{1}{2}\hbar \tag{1.14}$$

This relation is known as the Heisenberg uncertainty principle. According to quantum mechanics it is a fundamental property of nature that any attempt to make simultaneous measurements of position and momentum are subject to this limitation. It should be noted that these results are independent of the mass of the electron and therefore are not a result of the electron recoil. They would apply equally well to a particle of large mass, which would not move significantly as a result of photon scattering, although its *momentum* (mv where v is the velocity) uncertainty would still be given by (1.12). The Heisenberg uncertainty principle is more subtle than the popular idea of the value of one property being disturbed when the other is measured. We return to this point in our more general discussion of the uncertainty principle in chapter 4.

1.6 The rest of this book

In the next two chapters we discuss the nature and properties of matter waves in more detail and show how to obtain a wave equation whose solutions determine the energy levels of bound systems. We shall do this by considering one-dimensional waves in chapter 2, where we shall obtain qualitative agreement with experiment; in the following chapter we shall extend our treatment to three-dimensional systems and obtain excellent quantitative agreement between the theoretical results and experimental values of the energy levels of the hydrogen atom. At the same time we shall find that this treatment is incomplete and leaves many important questions unanswered. Accordingly, in chapter 4 we shall set up a more formal version of quantum mechanics within which the earlier results are included but which can, in principle, be applied to any physical system. This will prove to be a rather abstract process and prior familiarity with the results discussed in the earlier chapters will be a great advantage in understanding it. Having set up the general theory, it is then developed in subsequent chapters and discussed along with its applications to a number of problems such as the quantum theory of angular momentum and the special properties of systems containing a number of identical particles. Chapter 11 consists of an elementary introduction to relativistic quantum mechanics and quantum field theory, while chapter 12 discusses some examples of the applications of quantum mechanics to the processing of information that were developed towards the end of the twentieth century. The last chapter contains a detailed discussion of some of the conceptual problems of quantum mechanics. Chapters 7 to 13 are largely self-contained and can be read in a different order if desired.

Finally we should point out that photons, which have been referred to quite extensively in this chapter, will hardly be mentioned again except in

passing. This is primarily because a detailed treatment requires a discussion of the quantization of the electromagnetic field. We give a very brief introduction to quantum field theory in chapter 11, but anything more would require a degree of mathematical sophistication which is unsuitable for a book at this level. We shall instead concentrate on the many physical phenomena that can be understood by considering the mechanical system to be quantized and treating the electromagnetic fields semi-classically. However, it should be remembered that there are a number of important phenomena, particularly in high-energy physics, which clearly establish the quantum properties of electromagnetic waves, and field quantization is an essential tool in considering such topics.

Problems

1.1 The maximum energy of photoelectrons emitted from potassium is 2.1 eV when illuminated by light of wavelength 3×10^{-7} m and 0.5 eV when the light wavelength is 5×10^{-7} m. Use these results to obtain values for Planck's constant and the minimum energy needed to free an electron from potassium.

1.2 If the energy flux associated with a light beam of wavelength 3×10^{-7} m is 10 W m^{-2}, estimate how long it would take, classically, for sufficient energy to arrive at a potassium atom of radius 2×10^{-10} m in order that an electron be ejected. What would be the average emission rate of photoelectrons if such light fell on a piece of potassium 10^{-3} m^2 in area? Would you expect your answer to the latter question to be significantly affected by quantum-mechanical considerations?

1.3 An x-ray photon of wavelength 1.0×10^{-12} m is incident on a stationary electron. Calculate the wavelength of the scattered photon if it is detected at an angle of (i) 60°, (ii) 90° and (iii) 120° to the incident radiation.

1.4 A beam of neutrons with known momentum is diffracted by a single slit in a geometrical arrangement similar to that shown for the double slit in figure 1.2. Show that an approximate value of the component of momentum of the neutrons in a direction perpendicular to both the slit and the incident beam can be derived from the single-slit diffraction pattern. Show that the uncertainty in this momentum is related to the uncertainty in the position of the neutron passing through the slit in a manner consistent with the Heisenberg uncertainty principle. (This example is discussed in more detail in chapter 4.)

Chapter 2

The one-dimensional Schrödinger equations

In the previous chapter we have seen that electrons and other subatomic particles sometimes exhibit properties similar to those commonly associated with waves: for example, electrons of the appropriate energy are diffracted by crystals in a manner similar to that originally observed in the case of x-rays. We have also seen that the energy and momentum of a free particle can be expressed in terms of the angular frequency and wavevector of the associated plane wave by the de Broglie relations (1.10).

We are going to develop these ideas to see how the wave properties of the electrons bound within atoms can account for atomic properties such as line spectra. Clearly atoms are three-dimensional objects, so we shall eventually have to consider three-dimensional waves. However, this involves somewhat complex analysis, so in this chapter we shall begin by studying the properties of electron waves in one dimension.

In one dimension the wavevector and momentum of a particle can be treated as scalars so the de Broglie relations can be written as

$$E = \hbar\omega \qquad p = \hbar k \tag{2.1}$$

We shall use these and the properties of classical waves to set up a wave equation, known as the *Schrödinger wave equation*, appropriate to these 'matter waves'. When we solve this equation for the case of particles that are not free but move in a potential well, we shall find that solutions are only possible for particular discrete values of the total energy. We shall apply this theory to a number of examples and compare the resulting energy levels with experimental results.

2.1 The time-dependent Schrödinger equation

Consider a classical plane wave (such as a sound or light wave) moving along the x axis. Its displacement at the point x at time t is given by the real part of the complex quantity A where

$$A(x, t) = A_0 \exp[i(kx - \omega t)] \tag{2.2}$$

(In the case of electromagnetic waves, for example, the real part of A is the magnitude of the electric field vector.) This expression is the solution to a *wave equation* and the form of wave equation applicable to many classical waves is

$$\frac{\partial^2 A}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 A}{\partial t^2} \tag{2.3}$$

where c is a real constant equal to the wave velocity. If we substitute the right-hand side of (2.2) into (2.3), we see that the former is a solution to the latter if

$$-k^2 = -\omega^2/c^2$$

that is

$$\omega = c|k| \tag{2.4}$$

We can see immediately that the equation governing matter waves cannot have the form (2.3), because (2.4) combined with the de Broglie relations (2.1) gives the linear relation

$$E = cp \tag{2.5}$$

whereas for non-relativistic free particles the energy and momentum are known to obey the classical relation

$$E = p^2/2m \tag{2.6}$$

In the case of matter waves, therefore, we must look for a wave equation of a different kind from (2.3). However, because we know that plane waves are associated with free particles, expression (2.2) must also be a solution to this new equation.

If the equations (2.1) and (2.6) are to be satisfied simultaneously, it is necessary that the frequency of the wave be proportional to the square of the wavevector, rather than to its magnitude as in (2.4). This indicates that a suitable wave equation might involve differentiating twice with respect to x, as in (2.3), but only once with respect to t. Consider, therefore, the equation

$$\frac{\partial^2 \Psi}{\partial x^2} = \alpha \frac{\partial \Psi}{\partial t} \tag{2.7}$$

where α is a constant and $\Psi(x, t)$ is a quantity known as the *wavefunction* whose significance will be discussed shortly. If we now substitute a plane wave of the form (2.2) for Ψ we find that this is a solution to (2.7) if

$$-k^2 = -i\alpha\omega$$

We are therefore able to satisfy (2.1) and (2.6) by defining α such that

$$\alpha = -2mi/\hbar$$

Substituting this into (2.7) and rearranging slightly we obtain the wave equation for the matter waves associated with free particles as

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} \tag{2.8}$$

We can verify that this equation meets all the previous requirements by putting Ψ equal to a plane wave of the form (2.2) and using the de Broglie relations (2.1) to get

$$E\Psi = (p^2/2m)\Psi \tag{2.9}$$

as expected.

However, so far we have only found an equation which produces the correct results for a free particle, whereas we are looking for a more general theory to include the case of a particle moving under the influence of a potential, $V(x,t)$. The total energy E in this case is equal to the sum of the kinetic and potential energies which suggests a possible generalization of (2.9) to

$$E\Psi = (p^2/2m + V)\Psi$$

which in turn suggests that the wave equation (2.9) could be similarly generalized to give

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi \tag{2.10}$$

Equation (2.10) was first obtained by Erwin Schrödinger in 1926 and is known as the *one-dimensional time-dependent Schrödinger equation*; its further generalization to three-dimensional systems is quite straightforward and will be discussed in the following chapter. We shall shortly obtain solutions to this equation for various forms of the potential $V(x,t)$, but in the meantime we shall pause to consider the validity of the arguments used to obtain (2.10).

It is important to note that these arguments in no way constitute a rigorous derivation of a result from more basic premises: we started with a limited amount of experimental knowledge concerning the properties of free particles and their associated plane waves, and we ended up with an equation for the wavefunction associated with a particle moving under the influence of a general potential! Such a process whereby we proceed from a particular example to a more general law is known as *induction*, in contrast with *deduction* whereby a particular result is derived from a more general premise.

Induction is very important in science, and is an essential part of the process of the development of new theories, but it cannot by itself establish the truth of the general laws so obtained. These remain inspired guesses until physical properties have been deduced from them and found to be in agreement with the results of experimental measurement. Of course, if only one case of disagreement were to be found, the theory would be falsified and we should need to look for a more general law whose predictions would then have to agree with experiment in this

new area, as well as in the other cases where the earlier theory was successful. The Schrödinger equation, and the more general formulation of quantum mechanics to be discussed in chapter 4, have been set up as a result of the failure of classical physics to predict correctly the results of experiments on microscopic systems; they must be verified by testing their predictions of the properties of systems where classical mechanics has failed and also where it has succeeded. Much of the rest of this book will consist of a discussion of such predictions and we shall find that the theory is successful in every case; in fact the whole of atomic physics, solid state physics and chemistry obey the principles of quantum mechanics. The same is true of nuclear and particle physics, although an understanding of very high-energy phenomena requires an extension of the theory to include relativistic effects and field quantization, which are briefly discussed in chapter 11.

The wavefunction

We now discuss the significance of the wavefunction, $\Psi(x,t)$, which was introduced in equation (2.7). We first note that, unlike the classical wave displacement, the wavefunction is essentially a complex quantity. In the classical case the complex form of the classical wave is used for convenience, the physical significance being confined to its real part which is itself a solution to the classical wave equation. In contrast, neither the real nor the imaginary part of the wavefunction, but only the full complex expression, is a solution to the Schrödinger equation. It follows that the wavefunction cannot itself be identified with a single physical property of the system. However, it has an indirect significance which we shall now discuss—again using an inductive argument.

When we discussed diffraction in chapter 1, we saw that, although the behaviour of the individual particles is random and unpredictable, after a large number have passed through the apparatus a pattern is formed on the screen whose intensity distribution is proportional to the intensity of the associated wave. That is, the number of particles arriving at a particular point per unit time is proportional to the square of the amplitude of the wave at that point. It follows that if we apply these ideas to matter waves and consider one particle, the probability that it will be found in a particular place may well be proportional to the square of the modulus of the wavefunction there. Thus, if $P(x,t)dx$ is the probability that the particle is at a point between x and $x+dx$ at a time t, then $P(x,t)$ should be proportional to $|\Psi(x,t)|^2$. This means that, if we know the wavefunction associated with a physical system, we can calculate the probability of finding a particle in the vicinity of a particular point. This interpretation of the wavefunction was first suggested by Max Born and is known as the Born postulate. It is a fundamental principle of quantum mechanics that this probability distribution represents all that can be predicted about the particle position: in contrast to classical mechanics which assumes that the position of a particle is always known (or at least knowable) quantum mechanics states that it is almost always uncertain and indeterminate. We shall discuss this indeterminacy in more

detail in chapter 4, where we shall extend this argument to obtain expressions for the probability distributions governing the measurement of other physical properties, such as the particle momentum, and see how these ideas relate to the uncertainty principle. It is this 'probabilistic' aspect of quantum mechanics which has given rise to many of the conceptual difficulties associated with the subject, and we shall discuss some of these in chapter 13.

We can now impose an important constraint on the wavefunction: at any time we must certainly be able to find the particle somewhere, so the total probability of finding it with an x coordinate between plus and minus infinity must be unity. That is,

$$\int_{-\infty}^{\infty} P(x,t)\,dx = 1 \tag{2.11}$$

Now, referring back to (2.10), we see that if Ψ is a solution to the Schrödinger equation then $C\Psi$ is also a solution where C is any constant (a differential equation with these properties is said to be *linear*). The scale of the wavefunction can therefore always be chosen to ensure that the condition (2.11) holds and at the same time

$$P(x,t) = |\Psi(x,t)|^2 \tag{2.12}$$

This process is known as *normalization*, and a wavefunction which obeys these conditions is said to be *normalized*. The phase of C, however, is not determined by the normalizing process, and it turns out that a wavefunction can always be multiplied by a phase factor of the form $\exp(i\alpha)$, where α is an arbitrary, real constant, without affecting the values of any physically significant quantities.

2.2 The time-independent Schrödinger equation

We now consider the case where the potential, V, is not a function of time and where, according to classical mechanics, energy is conserved. Much of this book will relate to the quantum mechanics of such 'closed systems' and we shall discuss the more general problem of time dependence in detail only in chapter 8. If V is time independent we can apply the standard 'separation of variables' technique to the Schrödinger equation, putting

$$\Psi(x,t) = u(x)T(t) \tag{2.13}$$

Substituting (2.13) into (2.10) and dividing both sides by Ψ, we get

$$i\hbar\frac{1}{T}\frac{dT}{dt} = -\frac{1}{u}\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} + V(x) \tag{2.14}$$

Now the left-hand side of this equation is independent of x while the right-hand side is independent of t, but the equation must be valid for all values of x and t. This can be true only if both sides are equal to a constant which we call E. Thus

$$i\hbar\frac{dT}{dt} = ET \tag{2.15}$$

and

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} + V(x)u = Eu \tag{2.16}$$

Equation (2.15) can be solved immediately leading to

$$T = \exp(-iEt/\hbar) \tag{2.17}$$

while the solutions to (2.16) depend on the particular form of $V(x)$. Equation (2.16) is known as the one-dimensional *time-independent Schrödinger equation*. In the special case of a free particle, the origin of potential energy can be chosen so that $V(x) = 0$ and a solution to (2.16) is then

$$u = A\exp(ikx)$$

where $k = (2mE/\hbar^2)^{1/2}$ and A is a constant. Thus the wavefunction has the form

$$\psi = A\exp[i(kx - \omega t)] \tag{2.18}$$

where $\omega = E/\hbar$. This is just the same plane-wave form which we had originally in the case of a free particle (2.2)—provided that the constant E is interpreted as the total energy of the system, so our theory is self-consistent at least.[1]

In the case of any closed system, therefore, we can obtain solutions to the time-dependent Schrödinger equation corresponding to a given value of the energy of the system by solving the appropriate time-independent equation and multiplying the solution by the time-dependent phase factor (2.17). Provided the energy of the system is known and remains constant (and it is only this case which we shall be considering for the moment) the phase factor, T, has no physical significance. In particular, we note that the probability distribution, $|\Psi|^2$, is now identical to $|u|^2$, so that the normalization condition (2.11) becomes

$$\int_{-\infty}^{\infty} |u|^2\, dx = 1 \tag{2.19}$$

We shall shortly proceed to obtain solutions to the time-independent Schrödinger equation for a number of forms of the potential, $V(x)$, but before doing so we must establish some boundary conditions that have to be satisfied if the solutions to the Schrödinger equation are to represent physically acceptable wavefunctions.

2.3 Boundary conditions

Besides fulfilling the normalization condition, a solution to the time-independent Schrödinger equation must obey the following boundary conditions:

[1] There are particular difficulties associated with the normalization of a wavefunction which has a form such as (2.18) and these are discussed in detail in chapter 9.

1. The wavefunction must be a continuous, single-valued function of position and time.

   This boundary condition ensures that the probability of finding a particle in the vicinity of any point is unambiguously defined, rather than having two or more possible values—as would be the case if the probability distribution $|\Psi|^2$ were a many-valued function of x (such as $\sin^{-1} x$, for example) or had discontinuities. Although, strictly speaking, this argument only requires $|\Psi|^2$ to be single valued, imposing the condition on the wavefunction itself ensures the successful calculation of other physical quantities;[2] an example of this occurs in the discussion of spherically symmetric systems in chapter 3.
2. The integral of the squared modulus of the wavefunction over all values of x must be finite.

   In the absence of this boundary condition, the wavefunction clearly could not be normalized and the probabilistic interpretation would not be possible. We use this condition to reject as physically unrealistic, solutions to the Schrödinger equation that are zero everywhere or which diverge strongly to infinity at any point. A modification of this boundary condition and the procedure for normalizing the wavefunction is necessary in the case of free particles, and this is discussed in chapter 9.
3. The first derivative of the wavefunction with respect to x must be continuous everywhere except where there is an infinite discontinuity in the potential.

   This boundary condition follows from the fact that a finite discontinuity in $\partial\Psi/\partial x$ implies an infinite discontinuity in $\partial^2\Psi/\partial x^2$ and therefore, from the Schrödinger equation, in $V(x)$.

Having set up these boundary conditions we are now ready to consider the solutions to the Schrödinger equation in some particular cases.

2.4 Examples

(i) *The Infinite Square Well* As a first example we consider the problem of a particle in the potential $V(x)$ that is illustrated in figure 2.1 and is given by

$$V = 0 \qquad -a \leqslant x \leqslant a \tag{2.20}$$

$$V = \infty \qquad |x| > a \tag{2.21}$$

This is known as an *infinite square well*.

In the first region, the time-independent Schrödinger equation (2.16) becomes

$$\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} + Eu = 0 \tag{2.22}$$

[2] A detailed discussion of this point has been given by E. Merzbacher, *Am. J. Phys.*, vol. 30, p. 237, 1962.

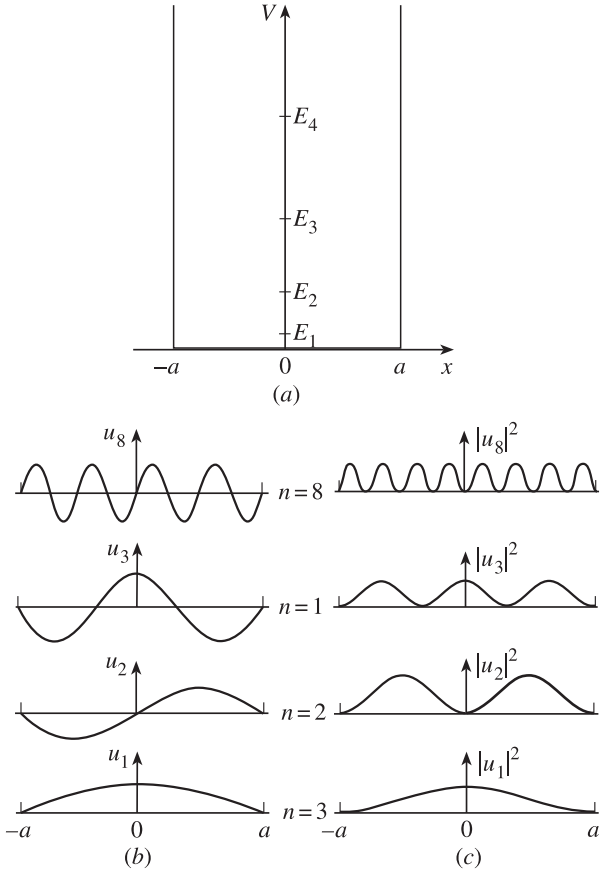


Figure 2.1. (*a*) shows the potential V as a function of x for an infinite square well, along with the energy levels of the four lowest energy states. The wavefunctions and position probability distributions corresponding to energy states with $n = 1, 2, 3$, and 8 are shown in (*b*) and (*c*) respectively.

The general solution to this equation is well known and can be verified by substitution. It can be written in the form

$$u = A \cos kx + B \sin kx \tag{2.23}$$

where A and B are constants and $k = (2mE/\hbar^2)^{1/2}$.

In the region outside the well where the potential is infinite, the Schrödinger equation can be satisfied only if the wavefunction is zero. We now apply the first boundary condition which requires the wavefunction to be continuous at $x = \pm a$

and therefore equal to zero at these points. Thus

$$\left.\begin{aligned} A\cos ka + B\sin ka &= 0 \\ A\cos ka - B\sin ka &= 0 \end{aligned}\right\} \quad \text{and} \tag{2.24}$$

Hence, either

$$\left.\begin{aligned} B = 0 \quad & \text{and} \quad \cos ka = 0 \\ & \text{that is,} \quad k = n\pi/2a \qquad n = 1, 3, 5, \ldots \\ A = 0 \quad & \text{and} \quad \sin ka = 0 \\ & \text{that is,} \quad k = n\pi/2a \qquad n = 2, 4, 6, \ldots \end{aligned}\right\} \quad \text{or} \tag{2.25}$$

These conditions, combined with the definition of k following (2.23), mean that solutions consistent with the boundary conditions exist only if

$$E \equiv E_n = \hbar^2\pi^2 n^2/8ma^2 \tag{2.26}$$

In other words, the energy is quantized. Application of the normalization condition (2.19) leads to the following expressions for the time-independent part of the wavefunction, which we now write as u_n:

$$\left.\begin{aligned} u_n &= a^{-1/2}\cos(n\pi x/2a) \qquad \text{for} \quad n \text{ odd} \\ u_n &= a^{-1/2}\sin(n\pi x/2a) \qquad \text{for} \quad n \text{ even} \end{aligned}\right\} \quad \text{if } -a \leqslant x \leqslant a$$
$$\text{and} \quad u_n = 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{if } |x| > a \tag{2.27}$$

These expressions are illustrated graphically in figure 2.1(*b*) for a number of values of n. We see that the wavefunction is either symmetric ($u_n(x) = u_n(-x)$) or antisymmetric ($u_n(x) = -u_n(-x)$) about the origin, depending on whether n is even or odd. This property is known as the *parity* of the wavefunction: symmetric wavefunctions are said to have even parity while antisymmetric wavefunctions are said to have odd parity. The possession of a particular parity is a general feature of the wavefunction associated with an energy state of a potential which is itself symmetric (i.e. when $V(x) = V(-x)$).

Remembering that the probability distribution for the particle position is given by $|u(x)|^2$, we see from figure 2.1 that, in the lowest energy state, the particle is most likely to be found near the centre of the box, while in the first excited state its most likely positions are near $x = \pm a/2$. For states of comparatively high energy, the probability distribution has the form of a large number of closely spaced oscillations of equal amplitude.

We can use these results to get some idea of how the Schrödinger equation can be used to explain atomic properties. The typical size of an atom is around 10^{-10} m and the mass of an electron is 9.1×10^{-31} kg. Taking the first of these to be a and substituting into (2.26) leads to the expression

$$E_n \simeq 1.5 \times 10^{-18} n^2 \text{ J}$$

The energy difference between the first and second levels is then 4.5×10^{-18} J (28 eV) so that a photon emitted in a transition between these levels would have a wavelength of about 4.4×10^{-8} m, which is of the same order as that observed in atomic transitions. If we perform a similar calculation with m the mass of a proton (1.7×10^{-27} kg) and a the order of the diameter of a typical nucleus (2×10^{-15} m) the energy difference between the first and second levels is now 5×10^{-12} J (34 MeV) which is in order-of-magnitude agreement with experimental measurements of nuclear binding energies. Of course, neither the atom nor the nucleus is a one-dimensional box, so we can only expect approximate agreement at this stage; quantitative calculations of atomic and nuclear energy levels must wait until we develop a full three-dimensional model in the next chapter.

One of the important requirements of a theory of microscopic systems is that it must produce the same results for macroscopic systems as are successfully predicted by classical mechanics. This is known as the *correspondence principle*. Applied to the present example, in the classical limit we expect no measurable quantization of the energy and a uniform probability distribution—because the particle is equally likely to be anywhere in the box. We consider a particle of mass 10^{-10} kg (e.g. a small grain of salt) confined to a box of half-width 10^{-6} m. These quantities are small on a macroscopic scale although large in atomic terms. The quantum states of this system then have energies

$$E_n = 1.4 \times 10^{-46} n^2 \text{ J}$$

The minimum energy such a system could possess would be that corresponding to the thermal energy associated with a single degree of freedom. Even at a temperature as low as 1 K this is of the order of 10^{-23} J leading to a value for n of around 3×10^{11}. The separation between adjacent energy levels would then be 8×10^{-35} J and an experiment of the accuracy required to detect any effects due to energy quantization would be completely impossible in practice. At this value of n the separation between adjacent peaks in the probability distribution would be 3×10^{-18} m and an experiment to define the position of the particle to this accuracy or better would be similarly impossible.[3] Thus to all intents and purposes, quantum and classical mechanics predict the same results—all positions within the well are equally likely and any value of the energy is allowed—and the correspondence principle is verified in this case.

(ii) *The Finite Square Well* We now consider the problem where the sides of the well are not infinite, but consist of finite steps. The potential, illustrated in

[3] If the energy of the system is not precisely defined then the exact value of n will be unknown. It will be shown later (chapter 4) that this implies that the wavefunction is then a linear combination of the wavefunctions of the states within the allowed energy span. The corresponding probability distribution is then very nearly uniform across the well—in even better agreement with the classical expectation.

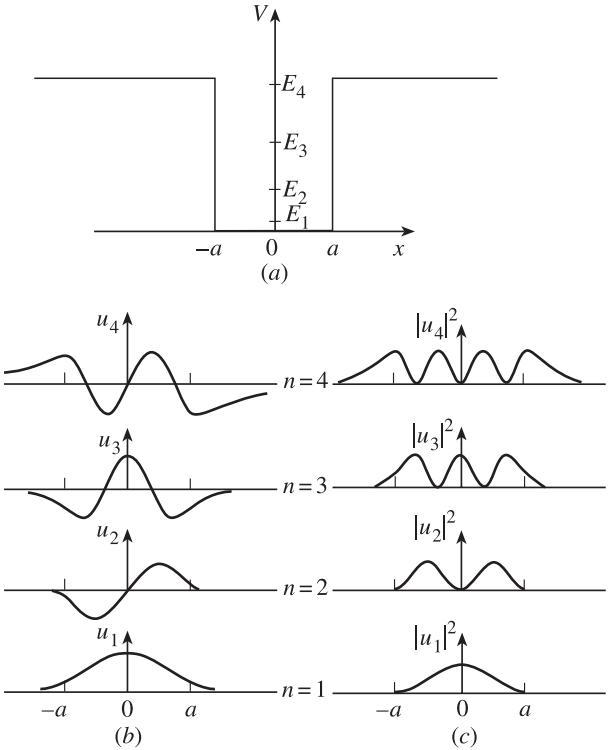


Figure 2.2. (*a*) shows the potential V as a function of x for a finite square well in the case where $V_0 = 25\hbar^2/2ma^2$, along with the energies of the four bound states. The wavefunctions and position probability distributions for these states are shown in (*b*) and (*c*) respectively.

figure 2.2, is then given by

$$\left.\begin{aligned} V &= 0 \qquad && -a \leqslant x \leqslant a \\ V &= V_0 && |x| > a \end{aligned}\right\} \tag{2.28}$$

We shall consider only bound states where the total energy E is less than V_0. The general solution to the Schrödinger equation in the first region is identical to the corresponding result in the infinite case (2.23). In the region $|x| > a$, however, the Schrödinger equation becomes

$$\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} - (V_0 - E)u = 0 \tag{2.29}$$

whose general solution is

$$u = C\exp(\kappa x) + D\exp(-\kappa x) \tag{2.30}$$

where C and D are constants and $\kappa = [2m(V_0 - E)/\hbar^2]^{1/2}$. We see at once that C must equal zero otherwise the wavefunction would tend to infinity as x tends to infinity in breach of the boundary conditions. Thus we have

$$u = D\exp(-\kappa x) \qquad x > a \tag{2.31}$$

A similar argument leads to

$$u = C\exp(\kappa x) \qquad x < -a \tag{2.32}$$

As the discontinuities in the potential at $x = \pm a$ are now finite rather than infinite, the boundary conditions require that both u and du/dx be continuous at these points. Thus we have, from (2.23), (2.31) and (2.32),

$$A\cos ka + B\sin ka = D\exp(-\kappa a) \tag{2.33}$$
$$-kA\sin ka + kB\cos ka = -\kappa D\exp(-\kappa a) \tag{2.34}$$
$$A\cos ka - B\sin ka = C\exp(-\kappa a) \tag{2.35}$$
$$kA\sin ka + kB\cos ka = \kappa C\exp(-\kappa a) \tag{2.36}$$

These equations lead directly to

$$2A\cos ka = (C + D)\exp(-\kappa a) \tag{2.37}$$
$$2kA\sin ka = \kappa(C + D)\exp(-\kappa a) \tag{2.38}$$
$$2B\sin ka = (D - C)\exp(-\kappa a) \tag{2.39}$$
$$2kB\cos ka = -\kappa(D - C)\exp(-\kappa a) \tag{2.40}$$

where (2.37) is obtained by adding (2.33) and (2.35), (2.38) is obtained by subtracting (2.34) from (2.36), and (2.39) and (2.40) are derived similarly. If we now divide (2.38) by (2.37) and (2.40) by (2.39) we get

$$\text{and} \quad \left.\begin{array}{llll} k\tan ka = \kappa & \text{unless } C = -D & \text{and} & A = 0 \\ k\cot ka = -\kappa & \text{unless } C = D & \text{and} & B = 0 \end{array}\right\} \tag{2.41}$$

The two conditions (2.41) must be satisfied simultaneously, so we have two sets of solutions subject to the following conditions:

$$\left.\begin{array}{llllll} \text{either} & k\tan ka = \kappa & C = D & \text{and} & B = 0 \\ \text{or} & k\cot ka = -\kappa & C = -D & \text{and} & A = 0 \end{array}\right\} \tag{2.42}$$

These, along with the definitions of k and κ, determine the energy levels and associated wavefunctions of the system.

Remembering that $k = (2mE)^{1/2}/\hbar$ and $\kappa = [2m(V_0 - E)]^{1/2}/\hbar$, we see that equations (2.42) determine the allowed values of the energy, just as the energy levels of the infinite well were determined by equations (2.25). However, in

the present case the solutions to the equations cannot be expressed as algebraic functions and we have to solve them numerically. One way of doing this is to use the definitions of k and κ to rewrite equations (2.42) as

$$\left.\begin{aligned} k^2a^2\tan^2(ka) &= (k_0^2 - k^2)a^2 \\ \text{and} \qquad k^2a^2\cot^2(ka) &= (k_0^2 - k^2)a^2 \end{aligned}\right\} \tag{2.43}$$

where $k_0^2 = 2mV_0/\hbar^2$. Equation (2.43) can be rewritten using standard trigonometric identities as

$$\left.\begin{aligned} ka &= n_1\pi + \cos^{-1}(ka/k_0a) \\ \text{and} \qquad ka &= n_2\pi - \sin^{-1}(ka/k_0a) \end{aligned}\right\} \tag{2.44}$$

where n_1 and n_2 are integers and the terms $n_1\pi$ and $n_2\pi$ are included because of the multivalued property of the inverse cosine and sine functions. In general, solutions will exist for several values of n_1 and n_2 corresponding to the different energy levels. However, it is clear that solutions do not exist if $n_1\pi$ or $n_2\pi$ is appreciably greater than k_0a because the arguments of the inverse cosine or sine would then have to be greater than one. This corresponds to the fact that there are a limited number of bound states with energies less than V_0.

Values for ka and hence E can be obtained by straightforward iteration. First, we evaluate k_0a from the values of V_0 and a for the particular problem. If we now guess a value for ka, we can substitute this into the right-hand side of one of (2.44) and obtain a new value of ka. This process usually converges to the correct value of ka. However, if the required value of ka is close to k_0a, iteration using (2.44) can fail to converge. Such cases can be successfully resolved by applying a similar iterative process to the equivalent equations

$$\left.\begin{aligned} ka &= k_0a\cos(ka - n_1\pi) \\ \text{and} \qquad ka &= k_0a\sin(n_2\pi - ka) \end{aligned}\right\} \tag{2.45}$$

The reader should try this for the case where $V_0 = 25\hbar^2/2ma^2$ so that k_0a equals 5.0. The ground state energy can be obtained from the first of (2.44) with $n_1 = 0$; starting with an initial value of ka anywhere between 1.0 and 2.0, ka should converge to 1.306 after a few iterations. If the exercise is repeated with $n_1 = 1$, another solution with $ka = 3.838$ should be obtained. However, if we try $n_1 = 2$, we are unable to obtain a solution, because the energy would now be greater than V_0. The remaining levels can be found by a very similar procedure using the second of equations (2.44) and (2.45). Table 2.1 sets out the details of all the possible solutions in this case, showing the energy levels both as fractions of V_0 and as fractions of the energies of the corresponding infinite-well states (2.26). The associated wavefunctions are shown in figure 2.2. Comparing these with the wavefunctions for the infinite square well (figure 2.1), we see that they are generally similar and, in particular, that they have a definite parity, being

Table 2.1. Values of the quantities ka, κa and E that are consistent with the boundary conditions for a potential well whose sides are of height V_0 when $V_0 = 25\hbar^2/2ma^2$. The energies of the corresponding states in the case where V_0 is infinite are represented by E_∞.

ka	κa	E/V_0	E/E_∞
1.306	4.826	0.069	0.691
2.596	4.273	0.270	0.682
3.838	3.205	0.590	0.663
4.907	0.960	0.964	0.610

either symmetric or antisymmetric about the point $x = 0$. However, one important difference between figures 2.2 and 2.1 is that in the former case the wavefunctions decay exponentially in the region $|x| > a$ instead of going to zero at $x = \pm a$. That is, the wavefunction penetrates a region where the total energy is less than V_0, implying that there is a probability of finding the particle in a place where it could not be classically as it would then have to have negative kinetic energy. This is another example of a quantum-mechanical result that is quite different from the classical expectation and we shall discuss it in more detail in the next section.

The penetration of the wavefunction into the classically forbidden region also results in the energy levels being lower than in the infinite square-well case (table 2.1) because the boundary conditions are now satisfied for smaller values of k. This effect is more noticeable for the higher energy levels and, conversely, we can conclude that in the case of a very deep well, the energy levels and wavefunctions of the low-lying states would be indistinguishable from those where V_0 was infinite. This point also follows directly from the boundary conditions: when $(V_0 - E)$ and therefore κ are very large, the conditions (2.42) become identical to (2.25).

2.5 Quantum mechanical tunnelling

We now turn to a more detailed discussion of effects associated with the penetration of the wavefunction into the classically forbidden region. Consider first a potential well bounded by barriers of finite height and width as in figure 2.3(a). As we have seen in the finite square well case, the wavefunction decays exponentially in the classically forbidden region and is still non-zero at the points $|x| = b$. In the regions where $|x| > b$, however, the total energy is again greater than the potential energy and the wavefunction is again oscillatory. It follows that there is a probability of finding the particle both inside and outside the potential well and also at all points within the barrier. Quantum mechanics therefore implies that a particle is able to pass through a potential energy barrier which, according to classical mechanics, should be impenetrable.

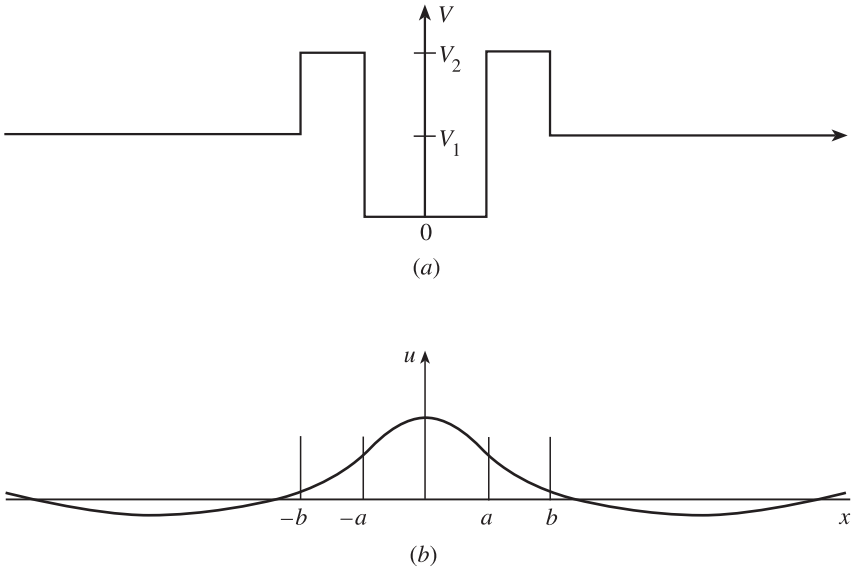


Figure 2.3. Particles in states with energy between 0 and V_0 can escape from the potential well illustrated in (*a*) by quantum-mechanical tunnelling. (*b*) shows the real part of the wavefunction of such a state.

This phenomenon is known as *quantum-mechanical tunnelling* or *the tunnel effect*.

To study the tunnel effect in more detail we consider the case of a beam of particles of momentum $\hbar k$ and energy $E = \hbar^2 k^2/2m$ approaching a barrier of height V_0 (where $V_0 > E$) and width b (see figure 2.4). A fraction of the particles will be reflected at the barrier with momentum $-\hbar k$, but some will tunnel through to emerge with momentum $\hbar k$ at the far side of the barrier. The incident, transmitted and reflected beams are all represented by plane waves, so the wavefunction on the incident side, which we take to be $x < 0$, is

$$u = A\exp(ikx) + B\exp(-ikx) \tag{2.46}$$

Inside the barrier the wavefunction has the same form as (2.30)

$$u = C\exp(\kappa x) + D\exp(-\kappa x) \tag{2.47}$$

and beyond the barrier, which is the region $x > b$, particles may emerge moving in the positive x direction, so the wavefunction will have the form

$$u = F\exp(ikx) \tag{2.48}$$

We note that because the barrier does not reach all the way to infinity, we cannot drop the first term in (2.47) as we did in the square-well case. The

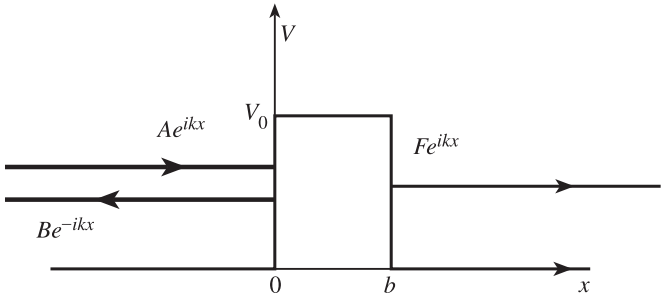


Figure 2.4. A beam of particles represented by a plane wave is incident on a potential barrier. Most particles are reflected, but some are transmitted by quantum-mechanical tunnelling.

boundary conditions requiring both u and du/dx be continuous at $x = 0$ and $x = b$ can be applied in much the same way as before.

$$\left.\begin{aligned} A + B &= C + D \\ A - B &= \frac{\kappa}{ik}(C - D) \\ C\exp(\kappa\, b) + D\exp(-\kappa b) &= F\exp(ikb) \\ C\exp(\kappa\, b) - D\exp(-\kappa b) &= \frac{ik}{\kappa}F\exp(ikb) \end{aligned}\right\} \tag{2.49}$$

Adding the first two equations and adding and subtracting the second two gives

$$\left.\begin{aligned} 2A &= \left(1 + \frac{\kappa}{ik}\right)C + \left(1 - \frac{\kappa}{ik}\right)D \\ 2C\exp(\kappa\, b) &= \left(1 + \frac{ik}{\kappa}\right)F\exp(ikb) \\ 2D\exp(-\kappa b) &= \left(1 - \frac{ik}{\kappa}\right)F\exp(ikb) \end{aligned}\right\} \tag{2.50}$$

We can combine these to express F in terms of A:

$$\frac{F}{A} = \frac{4i\kappa k}{(2i\kappa k + \kappa^2 - k^2)\exp(-\kappa b) + (2i\kappa k - \kappa^2 + k^2)\exp(\kappa b)}\exp(-ikb) \tag{2.51}$$

The fraction of particles transmitted is just the ratio of the probabilities of the particles being in the transmitted and incident beams, which is just $|F|^2/|A|^2$ and can be evaluated directly from (2.51). In nearly all practical cases, the tunnelling probability is quite small, so we can ignore the term in $\exp(-\kappa b)$ in the denominator of (2.51). In this case the tunnelling probability becomes

$$\frac{|F|^2}{|A|^2} = \frac{16\kappa^2k^2}{(\kappa^2 + k^2)^2}\exp(-2\kappa b) = \frac{16E(V_0 - E)}{V_0^2}\exp(-2\kappa b) \tag{2.52}$$

We see that this tunnelling probability is largely determined by the exponential decay of the wavefunction within the barrier: the lower and narrower the barrier is, the greater the likelihood of tunnelling. To apply this to the situation of tunnelling out of a well as in figure 2.3, we would first have to make a Fourier expansion of the wavefunction inside the well in terms of plane waves and then form an appropriately weighted sum of the transmission probabilities associated with them. In many cases, however, good semi-quantitative estimates can be made simply by considering the exponential decay of the wavefunction within the barrier.

A number of physical examples of tunnelling have been observed and two of these—alpha particle decay and cold electron emission—will now be described.

Alpha decay

It is well known that some nuclei decay radioactively emitting alpha particles. The alpha particle consists of two protons and two neutrons bound together so tightly that it can be considered as retaining this identity even when within the nucleus. The interaction between the alpha particle and the rest of the nucleus is made up of two components; the first results from the so-called strong nuclear force which is attractive, but of very short range, whereas the second is the Coulomb interaction which is repulsive (because both the alpha particle and the residual nucleus are positively charged) and acts at comparatively large distances. The total interaction potential energy is sketched in figure 2.5(a) as a function of the separation between the alpha particle and the nucleus, and we see that it is qualitatively similar to that shown in figure 2.3 and discussed previously. It follows that if the alpha particle occupies a quantum state whose energy is less than zero, it will remain there indefinitely and the nucleus will be stable. If, however, the form of the potential is such that the lowest energy state of the alpha particle is greater than zero, but less than V_0, it will be able to escape from the nucleus by quantum-mechanical tunnelling. The probability of such emission will depend on the actual shape of the barrier, particularly its height and width, which accounts for the large variation in the observed decay constants of different nuclei.

Cold electron emission

This phenomenon is observed when a strong electric field is directed towards the surface of a metal, resulting in the emission of electrons. This occurs even if the electrons are not thermally excited (which would be thermionic emission) and so do not have enough energy to escape classically. We first consider the situation in the absence of a field (figure 2.5(b)), when the electrons are confined within the metal by an energy barrier formed by the work function (see the discussion of the photoelectric effect in chapter 1). When the electric field is applied, the potential is changed, so that at a short distance from the surface of the metal the potential energy is less than the energy of the electrons inside the metal. Now, although

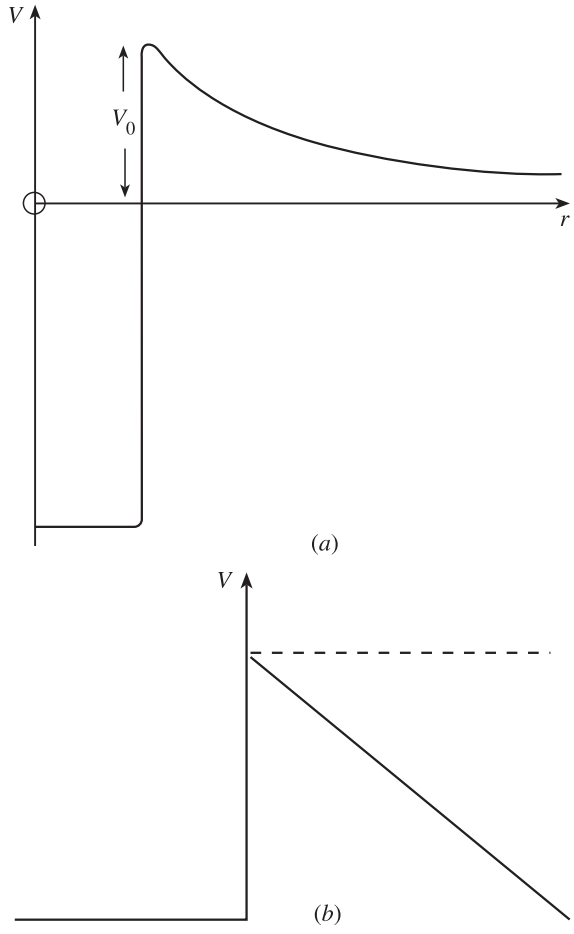


Figure 2.5. (*a*) shows the potential energy of interaction between an alpha particle and a nucleus as a function of its distance from the centre of the nucleus while (*b*) shows the potential energy of an electron near the surface of a metal with and without (broken line) an applied electric field. In each case the particles can pass through the potential barrier by quantum-mechanical tunnelling.

the electrons cannot classically penetrate the barrier at the metal surface, they can pass through by quantum-mechanical tunnelling and the observation of cold electron emission is therefore a confirmation of this effect.

In recent years, cold electron emission has been exploited in the *scanning tunnelling microscope*. In this instrument, an electric potential is maintained between a very sharp tungsten point and a metal surface above which it is held very closely. A tunnelling current between the surface and the point is

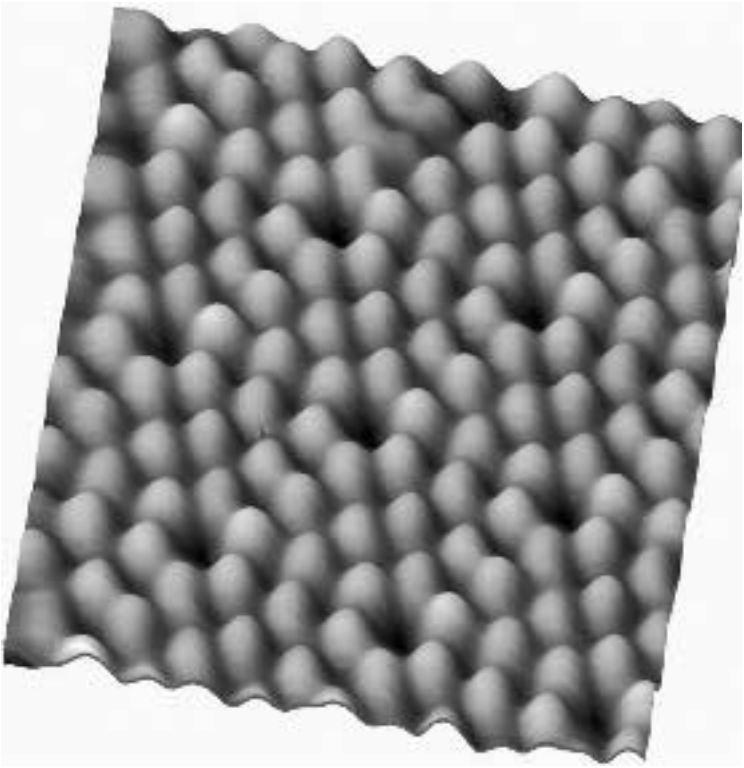


Figure 2.6. An image of the (111) surface of silicon obtained by scanning tunnelling microscopy. The bright peaks correspond to silicon atoms. The hexagonal symmetry is a characteristic feature of this surface. (Supplied by P. A. Sloan and R. E. Palmer of the Nanoscale Physics Research Laboratory in the University of Birmingham.)

measured and the point is scanned slowly across the metal surface. Variations in the tunnelling current then represent changes in the separation between the point and the source of electron emission. The method is very sensitive because of the exponential factor in (2.52). A typical value of κ is 10^{10} m^{-1}, so there are significant changes in the tunnelling current when the tip-to-sample distance changes by as little as 10^{-11} m. Using this technique, changes in the tunnelling current can be observed as the point moves over individual atoms, and so map the actual atomic structure on the metal surface. An example of this is shown in figure 2.6 which shows a silicon surface at atomic resolution. (In this case, in common with modern practice, a servo mechanism keeps the tunnelling current constant by moving the tip perpendicular to the surface, and the image is reconstructed from the record of the resulting tip movements.)

It should be noted that, although both these experiments imply that the alpha

particle or electron has passed through a classically forbidden region, in neither case has the particle been directly observed while undergoing this process. In fact such an observation appears always to be impossible; all known particle detectors are sensitive only to particles with positive kinetic energy so that, if we insert such a detector within the classically forbidden region, its presence implies that a 'hole' has been made in the potential so that the particle is no longer in such a region when it is detected.

2.6 The harmonic oscillator

We finish this chapter with a discussion of the energy levels and wavefunctions of a particle moving in a harmonic oscillator potential. This is an important example because many physical phenomena, including the internal vibrations of molecules and the motion of atoms in solids, can be described using it. It also provides an application of the method of series solution of differential equations, which is a technique that will be used again in the next chapter when we consider three-dimensional systems.

The harmonic oscillator potential has the form $V(x) = \frac{1}{2}Kx^2$ where K is a constant; classically a particle of mass m oscillates in this potential with an angular frequency $\omega_c = (K/m)^{1/2}$. The time-independent Schrödinger equation for this system can therefore be written in the form

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dx^2} + \frac{1}{2}m\omega_c^2 x^2 u = Eu \tag{2.53}$$

The subsequent mathematics is a good deal easier to follow if we first change variables from x to y, where $y = (m\omega_c/\hbar)^{1/2}x$ and define a constant $\alpha = (2E/\hbar\omega_c)$. Equation (2.53) now becomes

$$\frac{d^2u}{dy^2} + (\alpha - y^2)u = 0 \tag{2.54}$$

We first discuss the asymptotic form of solution in the region where y is very large so that the equation is approximately

$$\frac{d^2u}{dy^2} - y^2 u = 0 \tag{2.55}$$

We shall try as a solution to this equation

$$u = y^n \exp(-y^2/2) \tag{2.56}$$

where n is a positive integer. Differentiating twice with respect to y we get

$$\begin{aligned}\frac{d^2u}{dy^2} &= [n(n-1)y^{n-2} - (2n+1)y^n + y^{n+2}]\exp(-y^2/2) \\ &\simeq y^{n+2}\exp(-y^2/2) \quad \text{when } y \gg 1 \\ &= y^2 u\end{aligned}$$

Thus we see that (2.56) is the asymptotic form of the solution we are looking for, which suggests that a general solution to (2.54), valid for all values of y, might be

$$u(y) = H(y)\exp(-y^2/2) \tag{2.57}$$

where $H(y)$ is a function to be determined. Substituting from (2.57) into (2.54) we get

$$H'' - 2yH' + (\alpha - 1)H = 0 \tag{2.58}$$

where a prime indicates differentiation with respect to y.

We now write H in the form of a power series:

$$H = \sum_{p=0}^{\infty} a_p y^p \tag{2.59}$$

(Note that negative powers of y are not permitted as they produce physically unacceptable infinities at $y = 0$.) Hence

$$H' = \sum_{p=0}^{\infty} a_p p y^{p-1} \tag{2.60}$$

and

$$\begin{aligned} H'' &= \sum_{p=0}^{\infty} a_p p(p-1) y^{p-2} \\ &= \sum_{p=2}^{\infty} a_p p(p-1) y^{p-2} \end{aligned} \tag{2.61}$$

because the first two terms on the right-hand side of (2.61) vanish; thus,

$$H'' = \sum_{p'=0}^{\infty} a_{p'+2}(p'+2)(p'+1) y^{p'} \tag{2.62}$$

However, p' is just an index of summation so we can write it simply as p. Doing this and substituting from (2.59), (2.60) and (2.62) into (2.58) we get

$$\sum_{p=0}^{\infty} [(p+1)(p+2)a_{p+2} - (2p+1-\alpha)a_p] y^p = 0 \tag{2.63}$$

This can be true for all values of y only if the coefficient of each power of y vanishes, so we obtain the following recurrence relation:

$$\begin{aligned} a_{p+2}/a_p &= (2p+1-\alpha)/[(p+1)(p+2)] \\ &\to 2/p \qquad \text{as } p \to \infty \end{aligned} \tag{2.64}$$

This last expression is identical to the recurrence relation between successive terms of the power series for the function $\exp(y^2)(= \Sigma_n(y^{2n}/n!))$ so in general $H(y)$ will tend to infinity with y like $\exp(y^2)$, so that $u(y)$ diverges like $\exp(\frac{1}{2}y^2)$ leading to a physically unrealistic solution. This can be avoided only if the power series for H terminates after a finite number of terms. To obtain the conditions for such a termination, we first note that the series can be expressed as a sum of two series, one containing only even and the other only odd powers of y. By repeated application of (2.64) we can express the coefficients of y^p as functions of α multiplied by the constants a_0 or a_1 depending on whether p is even or odd, respectively. Either series, but not both, can be made to terminate by choosing α so that the numerator of (2.64) vanishes for some finite value of p (say $p = n$)—i.e. by putting α equal to $2n+1$. The other series cannot terminate simultaneously, but can be made to vanish completely if its leading coefficient (a_0 or a_1) is taken to be zero. Thus we have the following conditions for a physically acceptable solution to the Schrödinger equation in the case of a particle moving in a harmonic oscillator potential:

$$\left.\begin{array}{ll} \text{(i)} & \alpha = 2n+1 \qquad n = 0, 1, 2, \ldots \\ \text{(ii)} & a_1 = 0 \quad \text{if } n \text{ is even} \quad \text{and} \quad a_0 = 0 \quad \text{if } n \text{ is odd} \end{array}\right\} \tag{2.65}$$

If the first condition is combined with the definition of α, we find that the total energy of the system is quantized according to

$$E = E_n = (n + \tfrac{1}{2})\hbar\omega_c \tag{2.66}$$

Thus quantum mechanics predicts that the energy levels of a harmonic oscillator are equally spaced with an interval of $\hbar$ times the classical frequency and have a minimum value of $\frac{1}{2}\hbar\omega_c$ (known as the *zero-point energy*). Experimental confirmation of these results is obtained, for example, from observations of the properties of molecules. Thus a diatomic molecule can be considered as two point masses connected by a spring and this system can therefore undergo quantized oscillations. Photons absorbed and emitted by such a molecule then have frequencies that are multiples of the classical frequency of the oscillator. Thermal properties, like the heat capacity, of a gas composed of such molecules are also largely determined by energy quantization.

The polynomials $H_n(y)$ are known as Hermite polynomials. Expressions for them can be obtained by repeated application of (2.64) and the results of this procedure for the four lowest energy states are:

$$\left.\begin{aligned} H_0 &= 1 \\ H_1 &= 2y \\ H_2 &= 4y^2 - 2 \\ H_3 &= 8y^3 - 12y \end{aligned}\right\} \tag{2.67}$$

where the values of a_0 and a_1 have been chosen according to established convention. The wavefunctions $u_n(x)$ can now be obtained by multiplying $H_n(y)$ by the factor $\exp(-\frac{1}{2}y^2)$, making the substitution $y = (m\omega_c/\hbar)^{1/2}x$ and applying the normalization condition (2.19). The results of this procedure are shown in figure 2.7 where the wavefunctions are seen to have even or odd parity, depending on whether n is even or odd. We note that, like the wavefunctions associated with the energy states of a finite square well (figure 2.2), they are oscillatory within the classically permitted region and penetrate the classically forbidden region to some extent. Algebraic expressions for these wavefunctions are as follows:

$$\left.\begin{aligned} u_0 &= (m\omega_c/\pi\hbar)^{1/4}\exp(-m\omega_c x^2/2\hbar) \\ u_1 &= (4/\pi)^{1/4}(m\omega_c/\hbar)^{3/4}x\exp(-m\omega_c x^2/2\hbar) \\ u_2 &= (m\omega_c/4\pi\hbar)^{1/4}[2(m\omega_c/\hbar)x^2 - 1]\exp(-m\omega_c x^2/2\hbar) \\ u_3 &= (1/9\pi)^{1/4}(m\omega_c/\hbar)^{3/4}[(2m\omega_c/\hbar)x^2 - 3]x\exp(-m\omega_c x^2/2\hbar) \end{aligned}\right\} \tag{2.68}$$

The harmonic oscillator provides yet another example of the correspondence principle whereby the results of quantum mechanics tend to those of classical mechanics in the classical limit. To see this, we consider a particle of mass 10^{-10} kg oscillating with a classical angular frequency $\omega_c = 10^6$ rad s^{-1}. The separation of energy levels is then 10^{-28} J which corresponds to a temperature of about 10^{-5} K. At any normal temperature, therefore, the oscillator will be highly excited and the effects of energy quantization will be undetectable. We can also show that the probability distribution of the particle's position is similar to that expected classically if the oscillator is highly excited. Classically the position and velocity of an oscillating particle vary as

$$x = a\cos\omega_c t \qquad v = -a\omega_c \sin\omega_c t \tag{2.69}$$

where a is the amplitude of oscillation. Now an oscillating particle is more likely to be found near the limits of its oscillation where it is moving slowly than near the centre where it moves more quickly. Putting this point more quantitatively, the probability $P(x)dx$ that the particle be found in the region between x and dx is inversely proportional to the speed $v(x)$ at the point x. Hence using (2.69)

$$\begin{aligned} P(x) &= \frac{1}{\pi(a^2 - x^2)^{1/2}} \qquad |x| < a \\ &= 0 \qquad\qquad\qquad\quad |x| > a \end{aligned} \tag{2.70}$$

where the factor π ensures normalization. It is clear from figure 2.7(a) that the quantum-mechanical probability distribution is quite different from this for low energy states, but the agreement rapidly improves as n increases. A detailed comparison is shown for the case $n = 10$ in figure 2.7(b). This shows that the quantum-mechanical probability distribution is a rapidly oscillating function of x, but that if we average over these oscillations, the result is quite close to

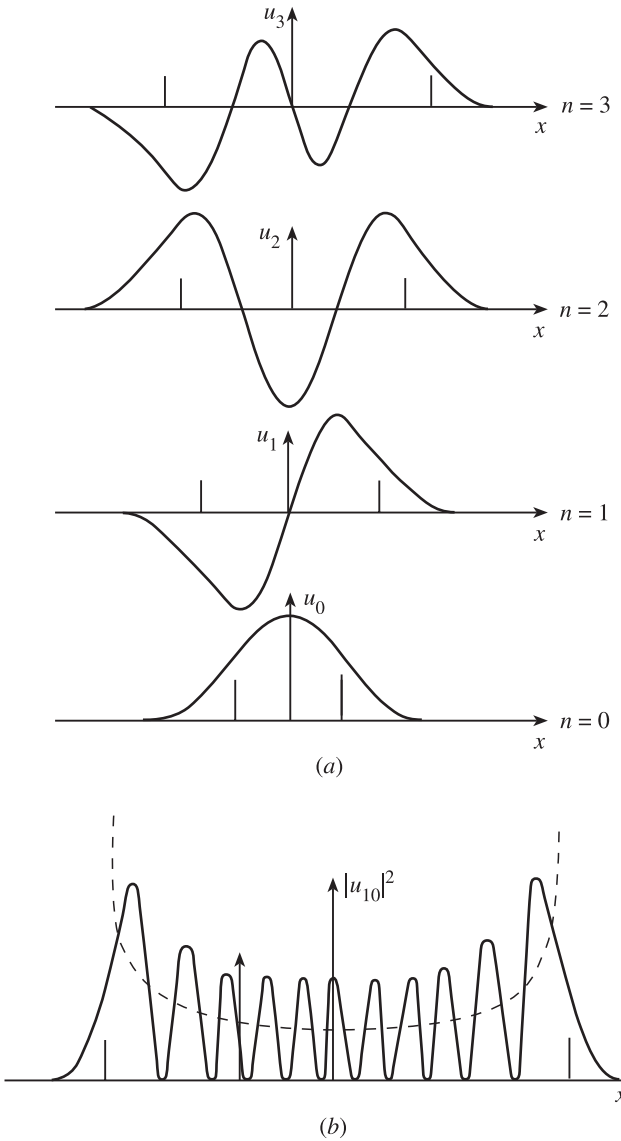


Figure 2.7. The wavefunctions corresponding to the four lowest energy states of the harmonic oscillator are shown in (*a*) while (*b*) shows the position probability distribution in the case where $n = 10$, compared with that calculated classically (broken curve). The limits of the classical oscillation are indicated by vertical lines in each case.

that given in (2.70). For larger values of n this agreement is even better and the separation of the adjacent maxima and minima in the quantum-mechanical probability distribution also becomes undetectably small, so that the predictions of quantum and classical mechanics are then indistinguishable.

Problems

2.1 An electron is confined to a one-dimensional potential well of width 6×10^{-10} m which has infinitely high sides. Calculate: (i) the three lowest allowed values of the electron energy; (ii) the wavelength of the electromagnetic wave that would cause the electron to be excited from the lowest to the highest of these three levels; (iii) all possible wavelengths of the radiation emitted following the excitation in (ii).

2.2 If u_m and u_n are the wavefunctions corresponding to two energy states of a particle confined to a one-dimensional box with infinite sides, show that

$$\int_{-\infty}^{\infty} u_n u_m \, dx = 0 \qquad \text{if } n \neq m$$

This is an example of 'orthogonality' which will be discussed in chapter 4.

2.3 Consider a particle of mass m subject to a one-dimensional potential $V(x)$ that is given by

$$V = \infty, \quad x < 0; \qquad V = 0, \quad 0 \leqslant x \leqslant a; \qquad V = V_0, \quad x > a$$

Show that bound ($E < V_0$) states of this system exist only if $k \cot ka = -\kappa$ where $k^2 = 2mE/\hbar^2$ and $\kappa^2 = 2m(V_0 - E)/\hbar^2$.

2.4 Show that if $V_0 = 9\hbar^2/2ma^2$, only one bound state of the system described in problem 2.3 exists. Calculate its energy as a fraction of V_0 and sketch its wavefunction, using an iteration similar to that discussed in Section 2.4.

2.5 Show that if $V(x) = V(-x)$, solutions to the time-dependent Schrödinger equation must have definite parity—that is, $u(x) = \pm u(-x)$.

Hint: Make the substitution $y = -x$ and show first that $u(x) = Au(-x)$ where A is a constant.

2.6 Consider a particle of mass m subject to the one-dimensional potential $V(x)$ that is given by

$$V = 0 \qquad \text{if } -a \leqslant x \leqslant a \qquad \text{or} \qquad \text{if } |x| > b$$
$$V = V_0 \qquad \text{if } a < |x| \leqslant b$$

where $b > a$. Write down the form of an even-parity solution to the Schrödinger equation in each region in the case where $E < V_0$. Note that the particle is not bound in this potential as there is always a probability of quantum-mechanical tunnelling, so solutions exist for all values of E. Show, however, that if $\kappa(b - a) \gg 1$ (where κ is defined as in problem 2.3) the probability of finding the particle inside the region $|x| < a$ is very small unless its energy is close to that of one of the bound states of a well of side $2a$ bounded by potential steps of height V_0. In the case where this condition is fulfilled exactly, obtain an expression for the ratio of the amplitudes of the wavefunction in the regions $|x| \leqslant a$ and $|x| > b$.

2.7 The hydrogen atom in a water molecule can vibrate in a direction along the O–H bond, and this motion can be excited by electromagnetic radiation of a wavelength about 4×10^{-6} m, but not by radiation of a longer wavelength. Calculate the effective spring constant for this vibration and the zero-point energy of the oscillator. Given that every molecular degree of freedom has a thermal energy of about $k_B T$, where k_B (Boltzmann's constant) $\simeq 1.4 \times 10^{-23}$ J K^{-1} and T is the temperature, what is the most probable vibrational state in the case of a water molecule in steam at 450 K?

2.8 Calculate the normalization constants for the two lowest energy states of a harmonic oscillator and verify that they are orthogonal in the sense defined in problem 2.2.

Chapter 3

The three-dimensional Schrödinger equations

In the previous chapter we saw that for a particle in a one-dimensional potential well, physically acceptable solutions to the one-dimensional Schrödinger equation are possible only for particular discrete values of the total energy. Moreover, in the case of an electron in a well of atomic dimensions, the spacings between these energy levels are in qualitative agreement with the separations experimentally observed in atoms. We also interpreted the square of the wavefunction as a probability distribution for the position of the particle. This led us to phenomena, such as quantum-mechanical tunnelling, that have been observed experimentally. The real world, however, is three-dimensional and, although one-dimensional examples often provide useful insights and analogies, we have to extend our theory into three dimensions before we can make quantitative predictions of most experimental results. In the present chapter, therefore, we shall set up the three-dimensional Schrödinger equation and obtain its solutions in a number of cases, culminating in a discussion of the hydrogen atom where we shall find that theory and experiment agree to a remarkable degree of accuracy.

3.1 The wave equations

In classical mechanics the total energy of a free particle of mass m and momentum p is given by

$$E = p^2/2m = (p_x^2 + p_y^2 + p_z^2)/2m \tag{3.1}$$

where p_x, p_y and p_z are the momentum components along the Cartesian axes x, y and z. The de Broglie relations (1.10) in three dimensions are:

$$\left.\begin{array}{ll} E = \hbar\omega & \\ \mathbf{p} = \hbar\mathbf{k} \qquad \text{that is} \qquad & p_x = \hbar k_x;\ p_y = \hbar k_y;\ p_z = \hbar k_z \end{array}\right\} \tag{3.2}$$

A wave equation whose solutions are consistent with these relations can be set up in exactly the same way as that described for the one-dimensional case—equations (2.1) to (2.8). We obtain

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2}\right) \tag{3.3}$$

where the wavefunction $\Psi(\mathbf{r}, t)$ is now a function of all three positional coordinates and the time. When the particle is subject to a potential $V(\mathbf{r}, t)$, equation (3.3) is generalized in the same way as in the one-dimensional case—equations (2.9) to (2.10)—to produce the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi \tag{3.4}$$

where we have used the vector operator ∇^2 (del-squared) which is defined so that

$$\nabla^2\Psi = \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2}$$

The generalization of the probabilistic interpretation of the wavefunction is also straightforward: if $P(\mathbf{r}, t)d\tau$ is the probability that the particle be found in the volume element $d\tau (\equiv dx\, dy\, dz)$ in the vicinity of the point $\mathbf{r}$ at time t, then

$$P(\mathbf{r}, t) = |\Psi(\mathbf{r}, t)|^2 \tag{3.5}$$

It follows directly that the normalization condition (2.12) becomes

$$\int |\Psi(\mathbf{r}, t)|^2\, d\tau = 1 \tag{3.6}$$

where the volume integral is performed over all space (as will always be assumed to be the case for volume integrals unless we specifically state otherwise).

When the potential V is independent of time, we can write $\Psi(\mathbf{r}, t) = u(\mathbf{r})T(t)$ and separate the variables to obtain the time-independent Schrödinger equation—cf. equations (2.13) to (2.17)

$$-\frac{\hbar^2}{2m}\nabla^2 u + Vu = Eu \tag{3.7}$$

along with

$$T(t) = \exp(-iEt/\hbar)$$

while the normalization condition (3.6) now becomes

$$\int |u(\mathbf{r})|^2\, d\tau = 1 \tag{3.8}$$

The boundary conditions on a wavefunction that is a solution to the time-independent Schrödinger equation also follow as a natural extension of the arguments set out in chapter 2: it must be a continuous, single-valued function of position and time; its squared modulus must be integrable over all space—which usually means that it must be everywhere finite; and its spatial derivatives ($\partial\Psi/\partial x$, $\partial\Psi/\partial y$ and $\partial\Psi/\partial z$) must all be continuous everywhere, except where there is an infinite discontinuity in V.

We shall now proceed to obtain solutions to the three-dimensional time-independent Schrödinger equation (3.7) in a number of particular cases. Unlike the one-dimensional case, the equation is now a partial differential equation which gives rise to considerable mathematical complications in general. We shall consider only those cases where the separation-of-variables technique can be used, but we shall still have to solve three ordinary differential equations in order to obtain a complete solution. We shall shortly consider spherically symmetric systems when we shall carry out this process in a spherical polar coordinate system, but we first discuss some simpler examples where the Schrödinger equation can be separated in Cartesian coordinates.

3.2 Separation in Cartesian coordinates

Consider the case where the potential $V(\mathbf{r})$ can be written as a sum of three quantities, each of which is a function of only one of the three Cartesian coordinates. That is,

$$V(\mathbf{r}) = V_1(x) + V_2(y) + V_3(z) \tag{3.9}$$

We now express the wavefunction as a product of three one-dimensional functions

$$u(\mathbf{r}) = X(x)Y(y)Z(z) \tag{3.10}$$

On substituting (3.9) and (3.10) into the time-independent Schrödinger equation (3.7), dividing through by u and rearranging slightly, we get

$$\left[-\frac{\hbar^2}{2m}\frac{1}{X}\frac{d^2X}{dx^2} + V_1(x)\right] + \left[-\frac{\hbar^2}{2m}\frac{1}{Y}\frac{d^2Y}{dy^2} + V_2(y)\right] + \left[-\frac{\hbar^2}{2m}\frac{1}{Z}\frac{d^2Z}{dz^2} + V_3(z)\right] = E \tag{3.11}$$

Each expression in square brackets is a function of only one of the variables (x, y, z), but the equation holds at all points in space. It follows that each of the expressions in square brackets must be equal to a constant and the sum of the

constants must equal E. Thus

$$\left.\begin{aligned} -\frac{\hbar^2}{2m}\frac{d^2X}{dx^2} + V_1X &= E_1X \\ -\frac{\hbar^2}{2m}\frac{d^2Y}{dy^2} + V_2Y &= E_2Y \\ -\frac{\hbar^2}{2m}\frac{d^2Z}{dz^2} + V_3Z &= E_3Z \end{aligned}\right\} \tag{3.12}$$

where $E_1 + E_2 + E_3 = E$. Each of these equations has the form of the one-dimensional Schrödinger equation (2.16) so, in suitable cases, we can carry over results directly from the previous chapter.

Example 3.1 The three-dimensional 'box' This example relates to a potential that is zero inside a rectangular region of sides $2a \times 2b \times 2c$ and infinite outside. That is, we are considering a potential of the form

$$\left.\begin{aligned} V(\mathbf{r}) &= 0 \qquad \text{if } -a \leqslant x \leqslant a, -b \leqslant y \leqslant b \text{ and } -c \leqslant z \leqslant c \\ V(\mathbf{r}) &= \infty \qquad \text{if } |x| > a, |y| > b, \text{ or } |z| > c \end{aligned}\right\} \tag{3.13}$$

Each of the three separated equations (3.12) is now equivalent to the Schrödinger equation for the one-dimensional infinite square well and the boundary conditions ($X = 0$ if $x = \pm a$, etc.) are also the same. It therefore follows directly from (2.26) that the energy levels are given by

$$E_{n_1n_2n_3} = \frac{\hbar^2\pi^2}{8m}\left(\frac{n_1^2}{a^2} + \frac{n_2^2}{b^2} + \frac{n_3^2}{c^2}\right) \tag{3.14}$$

where n_1, n_2 and n_3 are integers and the complete wavefunction is given—cf. (2.27)—by

$$u_{n_1n_2n_3} = (abc)^{-1/2}\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}\left(\frac{n_1\pi x}{2a}\right)\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}\left(\frac{n_2\pi y}{2b}\right)\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}\left(\frac{n_3\pi z}{2c}\right) \tag{3.15}$$

where the cosine applies if the integer in the following argument is odd, and the sine if it is even.

We see that three quantum numbers (n_1, n_2 and n_3) are needed to specify the energy and wavefunction in this example, compared with only one in the one-dimensional case. This is a general feature of three-dimensional bound systems.

It is interesting to consider the special case where two sides of the box are equal, as this illustrates some important features that arise when a three-dimensional potential has symmetry. Putting $a = b$, (3.14) becomes

$$E_{n_1n_2n_3} = \frac{\hbar^2\pi^2}{8m}\left(\frac{n_1^2 + n_2^2}{a^2} + \frac{n_3^2}{c^2}\right) \tag{3.16}$$

There are now, in general, several different combinations of n_1, n_2 and n_3 that have the same energy: e.g. the states ($n_1 = 2$, $n_2 = 1$, $n_3 = 1$) and ($n_1 = 1$, $n_2 = 2$, $n_3 = 1$) whose wavefunctions are

$$\left.\begin{aligned} u_{211} &= (a^2c)^{-1/2}\sin(\pi x/a)\cos(\pi y/2a)\cos(\pi z/2c) \\ u_{121} &= (a^2c)^{-1/2}\cos(\pi x/2a)\sin(\pi y/a)\cos(\pi z/2c) \end{aligned}\right\} \tag{3.17}$$

When two or more quantum states have the same value of the energy we say that they are *degenerate*. Very often the degeneracy is closely associated with the symmetry of the potential and we can illustrate this in the present case by considering the geometrical relationship between the two wavefunctions given in (3.17). Figure 3.1 shows contour diagrams of the section at $z = 0$ through each of these functions. These are clearly equivalent to each other, apart from their orientation in space: thus u_{121} can be transformed into u_{211} by rotating it through 90° about the z axis. As the potential has square symmetry in the xy plane, we should not expect such a rotation to result in any physical change of the system. However, the position probability distribution ($|u|^2$) corresponding to either of these states, does not have the expected symmetry. If the system is in the state with wavefunction u_{211}, for example, the probability of finding it near ($a/2$, 0, 0) is quite large, while that of finding it near (0, $a/2$, 0) is zero—even though, as far as the potential is concerned, these points are equivalent. To resolve this apparent paradox we must carefully consider what information a measurement of the energy gives us about a degenerate system. If only the energy has been measured, we can conclude that the wavefunction is one of the two forms given in (3.17), but we cannot tell which. In the absence of further information, we have to assume that either state is equally probable. It follows that the appropriate expression for the position probability distribution is the *average* of $|u_{211}|^2$ and $|u_{121}|^2$. This quantity clearly does have the same symmetry as the potential, as is also shown in figure 3.1. This argument relies on the states being degenerate, as otherwise we could tell from an energy measurement which one was occupied. This illustrates the close connection between degeneracy and symmetry. Moreover, the fact that the squared moduli of the wavefunctions associated with the individual states do not have a direct physical significance illustrates the general principle that to make sense of quantum mechanics, we should concentrate only on those results that can be measured and avoid drawing conclusions about apparent consequences which cannot be directly tested.

We shall return to the topic of degeneracy when we discuss spherically symmetric systems later in this chapter and we shall also consider the topic more formally in chapter 4.

Example 3.2 The three-dimensional harmonic oscillator Another example of a three-dimensional system, where the Schrödinger equation can be separated in Cartesian coordinates, is the case of a particle moving in the potential

$$V(\mathbf{r}) = \tfrac{1}{2}K_1x^2 + \tfrac{1}{2}K_2y^2 + \tfrac{1}{2}K_3z^2 \tag{3.18}$$

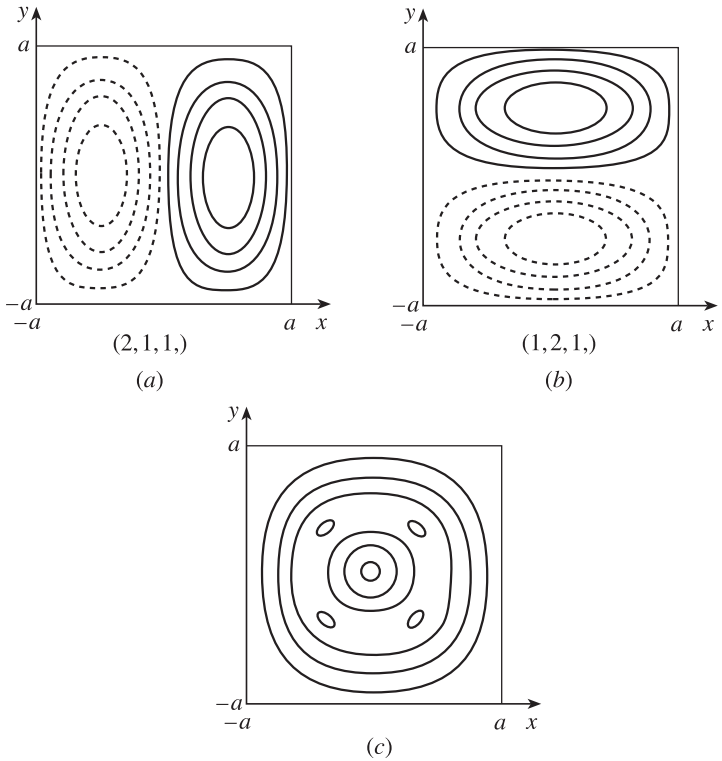


Figure 3.1. Sections at $z = 0$ through the three-dimensional wavefunctions of a particle in a rectangular box with $a = b$ corresponding to the states (*a*) $n_1 = 2, n_2 = 1, n_3 = 1$, and (*b*) $n_1 = 1, n_2 = 2, n_3 = 1$. Continuous lines represent positive contours and broken lines represent negative contours. A section at $z = 0$ through the position probability distribution calculated as the average of the squares of (*a*) and (*b*) is shown in (*c*). NB: This function is zero at the centre as well as the edges of the well and goes through a maximum in between.

The separated Schrödinger equations (3.12) are now

$$-\frac{\hbar^2}{2m}\frac{d^2X}{dx^2} + \frac{1}{2}m\omega_1^2 x^2 X = E_1 X \tag{3.19}$$

where $\omega_1 = (K_1/m)^{1/2}$, with similar equations for Y and Z. Each of these has the form of the one-dimensional harmonic oscillator equation (2.53) so we can use the results of this case (2.66) directly to get an expression for the energy levels of the three-dimensional oscillator

$$E_{n_1 n_2 n_3} = (n_1 + \tfrac{1}{2})\hbar\omega_1 + (n_2 + \tfrac{1}{2})\hbar\omega_2 + (n_3 + \tfrac{1}{2})\hbar\omega_3 \tag{3.20}$$

where n_1, n_2 and n_3 are positive integers. The wavefunctions also follow directly from the one-dimensional results:

$$u_{n_1 n_2 n_3} = H_{n_1}(x')H_{n_2}(y')H_{n_3}(z') \exp -\tfrac{1}{2}(x'^2 + y'^2 + z'^2) \tag{3.21}$$

where $x' = (m\omega_1/\hbar)^{1/2}x$, etc. and the H_{n_i}s are Hermite polynomials.

This example provides another illustration of how symmetry can result in degeneracy. If, for example, $K_1 = K_2 = K_3$ it follows from (3.18) that the potential is spherically symmetric and from (3.20) that all states with the same value of $(n_1 + n_2 + n_3)$ are degenerate.

3.3 Separation in spherical polar coordinates

Although there are applications of quantum mechanics in which Cartesian coordinates can be usefully employed, many interesting physical systems, particularly atoms and nuclei, are much more nearly spherical than they are rectangular. Spherically symmetric systems, where the potential $V(r)$ is independent of the direction of $\mathbf{r}$, are usually best treated using spherical polar coordinates (r, θ, ϕ). These are related to the Cartesian coordinates (x, y, z) by the expressions

$$\left.\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}\right\} \tag{3.22}$$

and the geometrical relationship between the two systems is shown in figure 3.2.

The Schrödinger equation (3.7) can be written in spherical polar coordinates as

$$-\frac{\hbar^2}{2m_e}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2}\right] + V(r)u = Eu \tag{3.23}$$

where the standard expression (derived in many mathematics textbooks) is used to express $\nabla^2 u$ in spherical polar coordinates, and we have represented the mass by m_e (a common symbol for electron mass) because m will be used later to represent a quantum number.

We now proceed to separate the variables and do so in two stages. We first put $u(r, \theta, \phi) = R(r)Y(\theta, \phi)$, substitute into (3.23), divide through by u and multiply through by r^2 to get

$$\left[-\frac{\hbar^2}{2m_e}\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + r^2V - r^2E\right] + \left[-\frac{\hbar^2}{2m_e}\frac{1}{Y}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) - \frac{\hbar^2}{2m_e}\frac{1}{Y}\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right] = 0 \tag{3.24}$$

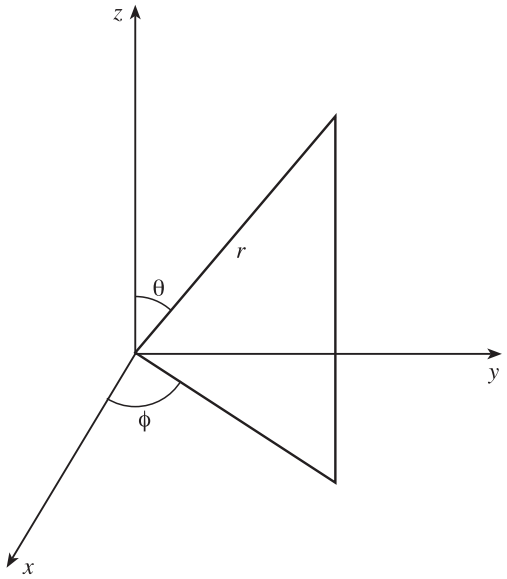


Figure 3.2. The geometrical relationship between the spherical polar coordinates r, θ, ϕ and the Cartesian axes x, y, z.

The contents of the first square bracket are independent of θ and ϕ and those of the second are independent of r, so they must be separately equal to constants, and the sum of the two constants must be equal to zero. We call these constants $-\lambda$ and λ, and obtain

$$-\frac{\hbar^2}{2m_e}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\left(V+\frac{\lambda}{r^2}\right)R=ER \tag{3.25}$$

and

$$-\frac{\hbar^2}{2m_e}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right)-\frac{\hbar^2}{2m_e}\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}=\lambda Y \tag{3.26}$$

Equation (3.26) does not contain the potential V. This means that if we can solve (3.26) for $Y(\theta,\phi)$, the solutions will represent the angular parts of the wavefunctions for *any* spherically symmetric potential $V(r)$ and we then 'only' have to solve the radial equation (3.25) to get the complete wavefunction in a particular case. We shall now show how the general solutions to (3.26) are obtained and return to the solution of the radial equation for particular potentials later.

Continuing the separation process, we put $Y(\theta,\phi)=\Theta(\theta)\Phi(\phi)$, substitute

into (3.18), multiply through by $\sin^2\theta/Y$ to get

$$\left[-\frac{\hbar^2}{2m_e}\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\lambda\sin^2\theta\right]+\left[-\frac{\hbar^2}{2m_e}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}\right]=0 \tag{3.27}$$

The contents of the first square bracket are independent of ϕ while those of the second are independent of θ, so they must each be equal to a constant and the sum of the two constants must be equal to zero. We call these constants $-\nu$ and ν, and get

$$-\frac{\hbar^2}{2m_e}\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\lambda\sin^2\theta\Theta+\nu\Theta=0 \tag{3.28}$$

and

$$-\frac{\hbar^2}{2m_e}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=\nu \tag{3.29}$$

The solution to (3.29) is straightforward, giving

$$\Phi=A\exp[\pm i(2m_e\nu/\hbar^2)^{1/2}\phi] \tag{3.30}$$

where A is a constant. We can now apply the condition that the wavefunction, and hence Φ, must be single valued[1] so that

$$\Phi(\phi+2\pi)=\Phi(\phi)$$

Thus

$$\exp[\pm i(2m_e\nu/\hbar^2)^{1/2}2\pi]=1$$

so that

$$(2m_e\nu/\hbar^2)^{1/2}=m \tag{3.31}$$

where m is an integer which can be positive or negative or zero. Substituting back into (3.30), we get

$$\Phi=(2\pi)^{-1/2}\exp(im\phi) \tag{3.32}$$

where the factor $(2\pi)^{-1/2}$ is included as a first step to normalizing the wavefunction; it ensures that

$$\int_0^{2\pi}|\Phi|^2\,d\phi=1 \tag{3.33}$$

We have now completed the solution of one of the three differential equations and obtained one quantum condition (3.31) along with one quantum number, m.

Returning to the equation for Θ (3.28), we can substitute from (3.31) and rearrange to get

$$\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+(\lambda'\sin^2\theta-m^2)\Theta=0 \tag{3.34}$$

[1] See footnote 2 in chapter 2.

where $\lambda' = 2m_e\lambda/\hbar^2$. The solution of this equation is made simpler if we make the substitution $v = \cos\theta$ and write $P(v) \equiv \Theta(\theta)$, leading to

$$\frac{d}{d\theta} = -\sin\theta\frac{d}{dv} = -(1-v^2)^{1/2}\frac{d}{dv}$$

Equation (3.34) then becomes

$$\frac{d}{dv}\left[(1-v^2)\frac{dP}{dv}\right] + \left[\lambda' - \frac{m^2}{1-v^2}\right]P = 0 \tag{3.35}$$

We first consider the simpler special case where m is equal to zero; equation (3.35) is then

$$\frac{d}{dv}\left[(1-v^2)\frac{dP}{dv}\right] + \lambda' P = 0 \tag{3.36}$$

The method of series solution which was previously employed in the case of the one-dimensional harmonic oscillator (section 2.6) can now be applied and we put

$$P = \sum_{p=0}^{\infty} a_p v^p \tag{3.37}$$

Hence

$$\begin{aligned}\frac{d}{dv}\left[(1-v^2)\frac{dP}{dv}\right] &= \frac{d}{dv}\sum_{p=0}^{\infty}[a_p p v^{p-1} - a_p p v^{p+1}] \\ &= \sum_{p=0}^{\infty} a_p p(p-1)v^{p-2} - \sum_{p=0}^{\infty} a_p p(p+1)v^p \\ &= \sum_{p'=0}^{\infty} a_{p'+2}(p'+2)(p'+1)v^{p'} - \sum_{p=0}^{\infty} a_p p(p+1)]v^p\end{aligned} \tag{3.38}$$

where $p' = p+2$; because the terms with $p = 0$ and $p = 1$ in the first summation on the second line are zero, the summation over p' in the last line begins at $p' = 0$. As p' is just an index of summation, we can re-write it as p and substitute from (3.38) into (3.36):

$$\sum_{p=0}^{\infty}\{a_{p+2}(p+2)(p+1) - a_p[p(p+1) - \lambda']\}v^p = 0$$

This can be true only if the coefficient of each power of v is zero, so we obtain the recurrence relation

$$\begin{aligned}\frac{a_{p+2}}{a_p} &= \frac{p(p+1) - \lambda'}{(p+1)(p+2)} \\ &\to 1 \qquad \text{as } p \to \infty\end{aligned} \tag{3.39}$$

Thus, for large p the series (3.37) is identical to the Taylor expansion of the function $(1-v)^{-1}$ which diverges to infinity at the point $v = 1$. Such a divergence in the wavefunction is not consistent with physical boundary conditions. It can be avoided if the series terminates at some finite value of p, say $p = l$, and if $a_0 = 0$ when l is odd and $a_1 = 0$ when l is even. We therefore obtain the second quantum condition

$$\lambda' = l(l+1) \tag{3.40}$$

where l is an integer which is greater than or equal to zero. Thus P, now written as P_l, is a polynomial of degree l which contains either only odd powers or only even powers of v. These polynomials are known as the Legendre polynomials and their properties are described in many mathematics textbooks. Explicit forms, corresponding to particular values of l, can be obtained from (3.40) and (3.39); for example

$$\left.\begin{aligned} P_0(v) &= 1 \\ P_1(v) &= v \\ P_2(v) &= \tfrac{1}{2}(3v^2 - 1) \\ P_3(v) &= \tfrac{1}{2}(5v^3 - 3v) \end{aligned}\right\} \tag{3.41}$$

where the values of the constants a_0 and a_1 have been chosen in accordance with established convention.

The solution of (3.35) in the general case of non-zero values of m is more complicated and the reader is referred to a mathematics textbook for the details. We note that (3.35) is independent of the sign of m, so we expect the solutions to be characterized by l and $|m|$ and we write them as $P_l^{|m|}(v)$. (Note: $|m|$ is *not* a power in this case.) It can be shown[2] that

$$P_l^{|m|}(v) = (1-v^2)^{|m|/2}\frac{d^{|m|}P_l}{dv^{|m|}} \tag{3.42}$$

We can use (3.42) to obtain a condition restricting the allowed values of m. P_l is a polynomial of degree l so its $|m|$th derivative, and hence $P_l^{|m|}$, will be zero if $|m|$ is greater than l. But if $P_l^{|m|}$ is zero, the whole wavefunction must be zero over all space, and this is physically unrealistic. We therefore have the condition

$$-l \leqslant m \leqslant l \tag{3.43}$$

We have now solved the differential equations in θ and ϕ so we can combine the solutions to obtain expressions for the angular part of the wavefunction, which we now write as $Y_{lm}(\theta, \phi)$, the suffixes l and m emphasizing the importance of

[2] The mathematically inclined reader can verify this result by substituting it into (3.35) and using Leibniz's expression for the nth derivative of a product to show that the result is equivalent to $(1-v^2)^{|m|/2}$ times the $|m|$th derivative of the left-hand side of equation (3.36).

these quantum numbers in characterizing the functions. We have

$$Y_{lm}(\theta,\phi) = (-1)^m \left[\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} P_l^{|m|}(\cos\theta)e^{im\phi} \qquad m \geqslant 0 \tag{3.44}$$

where it can be shown that the factor in square brackets ensures normalization of the function when it is integrated over all solid angles; that is

$$\int_0^{2\pi}\int_0^{\pi} |Y_{lm}(\theta,\phi)|^2 \sin\theta \, d\theta \, d\phi = 1 \tag{3.45}$$

The phase factors $(-1)^m$ in (3.44) are arbitrary, but chosen in accordance with established convention. The functions Y_{lm} are known as *spherical harmonics* and the reader is once again referred to an appropriate mathematics textbook for a discussion of their properties and a derivation of the form of the normalizing constant. Explicit expressions for the spherical harmonics with l less than or equal to two are given below and their shapes are illustrated in figure 3.3.

$$\left.\begin{aligned}
Y_{00} &= \frac{1}{(4\pi)^{1/2}} \\
Y_{10} &= \left(\frac{3}{4\pi}\right)^{1/2}\cos\theta \\
Y_{1\pm1} &= \mp\left(\frac{3}{8\pi}\right)^{1/2}\sin\theta e^{\pm i\phi} \\
Y_{20} &= \left(\frac{5}{16\pi}\right)^{1/2}(3\cos^2\theta - 1) \\
Y_{2\pm1} &= \mp\left(\frac{15}{8\pi}\right)^{1/2}\cos\theta\sin\theta e^{\pm i\phi} \\
Y_{2\pm2} &= \left(\frac{15}{32\pi}\right)^{1/2}\sin^2\theta e^{\pm 2i\phi}
\end{aligned}\right\} \tag{3.46}$$

A notable feature of figure 3.3 is that the wavefunctions have a particular orientation in space even though the potential is spherically symmetric and the direction of the z axis (sometimes known as the axis of quantization) is therefore arbitrary. This apparent contradiction is resolved in the same way as in the similar case of a particle in a square box discussed earlier. We first note that m does not enter equation (3.25) which determines the energy levels of the system, so there are always $2l + 1$ degenerate states that differ only in their values of m. If we measure the energy of such a system, we shall not be able to tell which of these wavefunctions is appropriate and we must therefore average their squared moduli in order to calculate the position probability distribution. The apparently

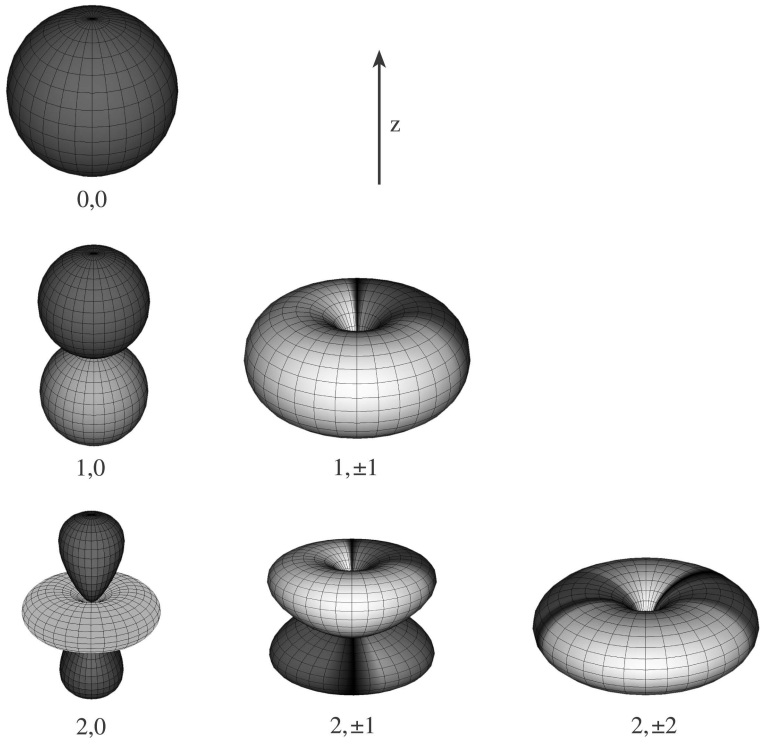


Figure 3.3. Representations of the shapes of the spherical harmonics with quantum numbers l, m, where $l \leqslant 2$ and the z axis is vertical. In the case of $m = 0$, the dark and light regions have opposite sign; when $m \neq 0$, the function is complex and its phase changes by $2m\pi$ during a complete circuit of the z axis.

angularly dependent part of this quantity will therefore be given by

$$(2l+1)^{-1} \sum_{m=-l}^{l} |Y_{lm}(\theta, \phi)|^2$$

It is one of the standard properties of the spherical harmonics that this quantity is spherically symmetric—as can be readily verified in the cases where $l = 0$, 1 and 2 by substituting the expressions given in equation (3.46)—so we once again see that the predictions of quantum mechanics concerning physically measurable quantities are consistent with what would be expected from the symmetry of the problem.

The physical significance of the quantum numbers l and m will be discussed in detail later (chapter 5). For the moment we note that m cannot be directly connected with the quantization of the energy of the system as this quantity

appears only in the radial equation, which we have yet to solve. It will turn out that l and m are associated with the quantization of the angular momentum of a particle in a central field: the square of the angular momentum has the value $l(l+1)\hbar^2$ and the z component of angular momentum has the value $m\hbar$.

The radial equation

We now turn our attention to the radial equation (3.25) which determines the energy levels of the system. Substituting the expression for λ obtained from the angular solution (3.40) and remembering that $\lambda' = 2m_e\lambda/\hbar^2$, we get

$$-\frac{\hbar^2}{2m_e}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\left[V(r)+\frac{l(l+1)\hbar^2}{2m_e r^2}\right]R = ER$$

This can be simplified by making the substitution $\chi(r) = rR(r)$ which gives

$$-\frac{\hbar^2}{2m_e}\frac{d^2\chi}{dr^2}+\left[V(r)+\frac{l(l+1)\hbar^2}{2m_e r^2}\right]\chi = E\chi \tag{3.47}$$

Apart from the second term within the square brackets, equation (3.47) is identical in form to the one-dimensional Schrödinger equation. However, an additional boundary condition applies in this case: χ must equal zero at $r = 0$ otherwise $R = r^{-1}\chi$ would be infinite at that point.[3]

As well as being mathematically convenient, the function $\chi(r)$ has a physical interpretation in that $|\chi|^2\,dr$ is the probability of finding the electron at a distance between r and $r + dr$ from the origin averaged over all directions. This follows from the fact that this probability is obtained by integrating $|\psi(r,\theta,\phi)|^2$ over a spherical shell of radius r and thickness dr. That is, it is given by

$$|R^2(r)|r^2dr\int_0^{2\pi}\int_0^{\pi}|Y(\theta,\phi)|^2\sin\theta\,d\theta\,d\phi = |\chi^2(r)|dr$$

using (3.45).

To progress further with the solution of the radial equation, the form of the potential $V(r)$ must be known, and in the next section we shall consider the particular example of the hydrogenic atom.

[3] The observant reader will have noticed that if $R \sim r^{-1}$ then $\int_0^r |R|^2r^2\,dr$ will be finite and therefore the boundary condition set out in chapter 2 is not breached. It can be shown that the condition $\chi = 0$ at $r = 0$ follows from the requirement that the solutions of the Schrödinger equation expressed in spherical polar coordinates must also be solutions when the equation is written in Cartesian coordinates. Further details on this point can be found in P. A. M. Dirac *The Principles of Quantum Mechanics* (Oxford 1974) chapter 6.

3.4 The hydrogenic atom

We are now ready to apply quantum theory to the real physical situation of an electron moving under the influence of a positively charged nucleus. If this nucleus consists of a single proton, the system is a hydrogen atom, but the theory is also applicable to the more general case of an atom with atomic number Z (and hence nuclear charge Ze) with all but one of its electrons removed (for example, He^+, Li^{2+}, etc.). In general, such a system is described as a *hydrogenic atom*. The potential energy of interaction between the electron and the nucleus is $-Ze^2/4\pi\varepsilon_0 r$, so equation (3.47) becomes[4] in this case

$$-\frac{\hbar^2}{2m_e}\frac{d^2\chi}{dr^2}+\left[-\frac{Ze^2}{4\pi\varepsilon_0 r}+\frac{l(l+1)\hbar^2}{2m_e r^2}\right]\chi=E\chi \tag{3.48}$$

The solution of equation (3.48) will again involve considerable manipulation which is simplified by making a suitable substitution. We define a new variable ρ so that

$$\rho=(-8m_eE/\hbar^2)^{1/2}r \tag{3.49}$$

(note that E is negative for bound states as the potential is zero when r is infinite) and hence

$$\frac{d^2\chi}{dr^2}=-\frac{8m_eE}{\hbar^2}\frac{d^2\chi}{d\rho^2} \tag{3.50}$$

Equation (3.48) now becomes

$$\frac{d^2\chi}{d\rho^2}-l(l+1)\frac{\chi}{\rho^2}+\left(\frac{\beta}{\rho}-\frac{1}{4}\right)\chi=0 \tag{3.51}$$

where the constant β is defined as

$$\beta=\left(-\frac{m_e}{2E}\right)^{1/2}\frac{Ze^2}{4\pi\varepsilon_0\hbar} \tag{3.52}$$

We first consider the solution to (3.51) in the case of very large ρ when the equation becomes

$$\frac{d^2\chi}{d\rho^2}-\frac{1}{4}\chi=0 \tag{3.53}$$

[4] We have assumed here that the electron is moving in the field of a fixed nucleus, but this will not be exactly true as the nucleus is also moving in the field of the electron. As is shown in chapter 10, this nuclear motion can be allowed for in exactly the same way as in classical mechanics by taking r to be the distance between the nucleus and the electron, and μ to be the reduced mass of the nucleus (mass m_N) and the electron (mass m_e). That is,

$$\mu=m_Nm_e/(m_N+m_e)$$

Because the mass of the electron is much smaller than that of the nucleus, μ is very nearly equal to m_e and the effect of nuclear motion is small. However, rather than using the reduced mass in the theoretical derivation, when comparison is made with experiment (such as the Rydberg constant discussed later) the experimental values have been adjusted to remove the effect of nuclear motion.

leading to

$$\chi \sim \exp(-\rho/2) \tag{3.54}$$

(where we have rejected a possible solution with positive exponent because it diverges to infinity at large ρ). This suggests that we try

$$\chi = F(\rho)\exp(-\rho/2) \tag{3.55}$$

as a solution to (3.51). On substitution we get

$$\frac{d^2F}{d\rho^2} - \frac{dF}{d\rho} - \frac{l(l+1)}{\rho^2}F + \frac{\beta}{\rho}F = 0 \tag{3.56}$$

We now look for a series solution to (3.56) and put

$$F = \sum_{p=1}^{\infty} a_p \rho^p \tag{3.57}$$

The lower limit of this summation is $p = 1$ rather than $p = 0$, otherwise F and, therefore, χ would not be zero at $\rho = 0$. Thus

$$\frac{dF}{d\rho} = \sum_{p=1}^{\infty} p a_p \rho^{p-1} \tag{3.58}$$

and

$$\begin{aligned}\frac{d^2F}{d\rho^2} &= \sum_{p=1}^{\infty} p(p-1)a_p\rho^{p-2} \\ &= \sum_{p=1}^{\infty}(p+1)pa_{p+1}\rho^{p-1}\end{aligned} \tag{3.59}$$

Also

$$\begin{aligned}F/\rho^2 &= \sum_{p=1}^{\infty} a_p \rho^{p-2} \\ &= a_1\rho^{-1} + \sum_{p=1}^{\infty} a_{p+1}\rho^{p-1}\end{aligned} \tag{3.60}$$

Substituting from equations (3.57) to (3.60) into (3.56) we get

$$\begin{aligned}-l(l+1)a_1\rho^{-1} + \sum_{p=1}^{\infty}[(p+1)pa_{p+1} - pa_p \\ - l(l+1)a_{p+1} + \beta a_p]\rho^{p-1} = 0\end{aligned} \tag{3.61}$$

The coefficient of each power of ρ must vanish so we have

$$a_1 = 0 \qquad \text{unless } l = 0$$

and

$$\frac{a_{p+1}}{a_p} = \frac{p - \beta}{p(p+1) - l(l+1)} \tag{3.62}$$

$$\to p^{-1} \qquad \text{as } p \to \infty \tag{3.63}$$

We first note that the denominator on the right-hand side of (3.62) is zero if $p = l$. This implies that a_{l+1} (and, by implication, all other a_p where p is greater than l) must be infinite unless a_l is zero. But if a_l equals zero, it follows from (3.62) that a_{l-1}, a_{l-2} etc. must also equal zero. We conclude, therefore, that all a_p with p less than or equal to l must be zero if the solution is to represent a physically realistic wavefunction. We also see that (3.63) is identical to the recurrence relation for the terms in the series expansion of $\exp(\rho)$ and so χ, which equals $F\exp(-\rho/2)$, will diverge like $\exp(\rho/2)$ as ρ tends to infinity. However, just as in the case of solutions to the harmonic oscillator and Legendre polynomial equations, this divergence can be prevented by ensuring that the series terminates after a finite number of terms. For this to occur at the term $p = n$ we must have

$$\beta = n > l \tag{3.64}$$

and hence, using (3.52)

$$E \equiv E_n = -\frac{m_e Z^2 e^4}{2(4\pi\varepsilon_0)^2 \hbar^2 n^2} \tag{3.65}$$

We have thus derived expressions for the discrete energy levels of the hydrogenic atom in terms of the mass of the electron, the nuclear charge and the fundamental constants e, $\hbar$ and ε_0. It should be noted that the energy levels (3.65) are not only independent of m, as would be expected from the earlier discussion, but are also independent of l. This additional degeneracy is a particular feature of the Coulomb potential and is not a general property of a spherically symmetric system.

It is now an acid test of the theory developed so far that we compare these energy levels with those experimentally measured from observations of atomic spectra. We saw in chapter 1 that the line spectra of hydrogen could be accounted for if the hydrogen atom were assumed to have a set of energy levels given by

$$E_n = -2\pi\hbar c R_0 / n^2 \tag{3.66}$$

where n is a positive integer and R_0 is the Rydberg constant. Comparison of (3.65) and (3.66) shows at once that these have the same form so that there is at least

qualitative agreement between theory and experiment. Quantitative comparison is made using the measured values of the fundamental constants

$$\begin{aligned} m_e &= 9.109\,381\,88 \times 10^{-31}\ \text{kg} \\ \varepsilon_0 &= 8.854\,187\,817 \times 10^{-12}\ \text{F m}^{-1} \\ \hbar &= 1.054\,571\,596 \times 10^{-34}\ \text{J s} \\ e &= 1.602\,176\,462 \times 10^{-19}\ \text{C} \\ c &= 2.997\,924\,58 \times 10^{8}\ \text{m s}^{-1} \end{aligned}$$

These lead to an estimate of R_0 of $10\,973\,731.6\ \text{m}^{-1}$, which is within one part in 10^{10} of the accepted best value.[5] Moreover, the tiny discrepancy is less than the estimated errors on the measurements of the relevant quantities. Similar agreement is obtained for other hydrogenic atoms when the appropriate values of the nuclear charge are substituted into equation (3.65). These results therefore represent an important test of quantum-mechanical theory, which it has passed with flying colours.[6] Our belief in quantum mechanics does not of course rest on this result alone: indeed an expression identical to (3.65) was derived by Niels Bohr using an earlier theory which was subsequently shown to be incorrect when applied to other more complex systems. However, although we shall compare the results of calculation and experiment on a number of other occasions when we shall always find agreement within the limits of experimental error, there are very few examples of physical quantities whose values can be both measured experimentally to such high precision and calculated exactly by solving the appropriate quantum-mechanical equations.

The hydrogenic atom wavefunctions

We now complete our consideration of the hydrogenic atom by discussing the form of the wavefunctions associated with the different energy levels. We previously saw that the radial part of the wavefunction is consistent with the boundary conditions only if the series (3.57) for F starts at the term $p = l + 1$ and terminates at $p = n$. We thus have

$$F_n(\rho) = \sum_{p=l+1}^{n} a_p \rho^p \tag{3.67}$$

where the coefficients a_p can be expressed in terms of a_{l+1} using the recurrence relation (3.62) with $\beta = n$. The results are known as the *Laguerre polynomials*.

[5] This value has been adjusted to remove the effects of nuclear motion (see footnote 4) and relativistic corrections (see chapter 11).

[6] Of course the quantum theory of atomic spectra is now so well established that formulae such as (3.65) are themselves used in determining the best values of the fundamental constants, but the fact that a wide variety of experimental data can be successfully and consistently used in this way is itself a confirmation of the theory.

We can then use (3.55) and the definition of ρ in terms of r to obtain $\chi(r)$ and hence $R(r)$. This can be combined with the appropriate spherical harmonic to produce an expression for the complete time-independent part of the wavefunction, $u(r, \theta, \phi)$. This will be normalized if the spherical harmonic has been normalized in accordance with (3.45) and the constant a_{l+1} has been chosen so that

$$\int_0^\infty |R|^2 r^2 \, dr = 1 \tag{3.68}$$

Formally, then, we have

$$u_{nlm} = R_{nl}(r) Y_{lm}(\theta, \phi) \tag{3.69}$$

where the suffixes indicate the dependence of the various functions on the quantum numbers n, l, and m. The wavefunctions corresponding to the five states of lowest energy as determined in this way are

$$\left.\begin{aligned} u_{100} &= (Z^3/\pi a_0^3)^{1/2} \exp(-Zr/a_0) \\ u_{200} &= (Z^3/8\pi a_0^3)^{1/2} (1 - Zr/2a_0) \exp(-Zr/2a_0) \\ u_{210} &= (Z^3/32\pi a_0^3)^{1/2} (Zr/a_0) \cos\theta \exp(-Zr/2a_0) \\ u_{21\pm1} &= \mp (Z^3/\pi a_0^3)^{1/2} (Zr/8a_0) \sin\theta \exp(\pm i\phi) \exp(-Zr/2a_0) \end{aligned}\right\} \tag{3.70}$$

where the constant a_0 is defined as

$$\begin{aligned} a_0 &= 4\pi\varepsilon_0 \hbar^2 / m_e e^2 \\ &= 0.529\,176\,6 \times 10^{-10} \text{ m} \end{aligned} \tag{3.71}$$

and is known as the *Bohr radius*.

The value of the azimuthal quantum number l is often denoted by a particular letter code: states with $l = 0$, 1, 2, and 3 are labelled s, p, d, and f respectively. This letter is sometimes prefixed by a number equal to the quantum number n, so that the first state in (3.70) is known as the $1s$ state, the second is $2s$ and the others are $2p$ states.

The radial parts of the wavefunctions (3.70) are plotted as functions of r in figure 3.4 for the case of the hydrogen atom where $Z = 1$. We see that the constant a_0 characterizes the width of the wavefunction of the lowest energy state and that this width increases for states of higher energy. We can combine (3.71) and (3.65) to express the energy levels in terms of a_0:

$$E_n = -\frac{Z^2 e^2}{2(4\pi\varepsilon_0) a_0 n^2} \tag{3.72}$$

As the potential energy is given by $V = -Ze^2/4\pi\varepsilon_0 r$, an electron with this total energy could only have positive kinetic energy for values of r less than $2n^2 a_0/Z$. We see in figure 3.4 that the exponential tails of the wavefunctions penetrate the

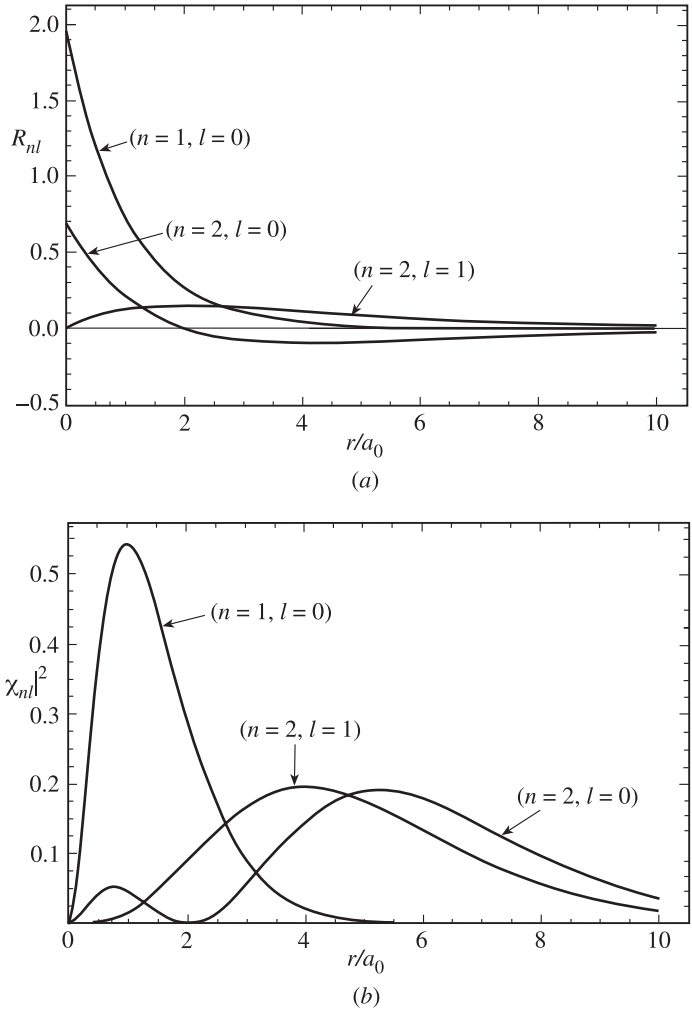


Figure 3.4. The radial parts, $R_{n,l}$, of the wavefunctions corresponding to some of the energy states of the hydrogen atom are shown in (*a*). The corresponding radial probability distributions, $|\chi_{n,l}|^2 = r^2|R_{n,l}|^2$, are displayed in (*b*).

classically forbidden region where r is greater than this in a manner very similar to that discussed in the one-dimensional cases in chapter 2.

Figure 3.4 also shows $|\chi^2(r)| = r^2|R^2(r)|$ for each state as a function of r. As we pointed out earlier, this expression equals the probability that the electron be found at a distance between r and $r + dr$ from the origin (in any direction). We see that this probability reaches a maximum at $r = a_0$ in the case of the ground-

state wavefunction. We particularly note that in all cases $|\chi|^2$ equals zero at the origin, even though the square of the wavefunction can have its largest value at that point. The reader should think carefully about this apparent contradiction and how it can be resolved by understanding the different nature of the two probability distributions represented by $|\psi|^2$ and $|\chi|^2$.

Problems

3.1 Calculate the energy levels and obtain expressions for the associated wavefunctions in the case of a particle moving in two dimensions in a rectangular, infinite-walled box of sides a and b. Discuss the degeneracy of the system and the symmetry of the position probability distribution when $a = b$.

3.2 What is the symmetry of the position probability distribution and how is it related to the degeneracy in the case of a particle confined to a three-dimensional box with cubic symmetry (that is, with $a = b = c$)?

3.3 A particle moves in two dimensions in a circularly symmetry potential. Show that the time-independent Schrödinger equation can be separated in plane polar coordinates and that the angular part of the wavefunction has the form $(2\pi)^{1/2}\exp(im\phi)$ where m is an integer. What is the symmetry of the position probability distribution in this case?

3.4 Consider a circularly symmetric two-dimensional system similar to that described in problem 3.3 where the potential is zero for all values of r less than a and infinite otherwise. Show that the radial part $R(r)$ of the wavefunction must satisfy the equation

$$\frac{d^2R}{d\rho^2} + \frac{1}{\rho}\frac{dR}{d\rho} + \left(1 - \frac{m^2}{\rho^2}\right)R = 0$$

where $\rho = (2m_eE/\hbar^2)^{1/2}r$. In the case where $m = 0$ show that $R = \sum_{k=0}^{\infty} A_k\rho^k$ where $A_k = 0$ if k is odd and $A_{k+2} = -A_k/(k+2)^2$. Given that the first zero of this function is at $\rho = 2.405$, obtain an expression for the energy of the ground state of the system.

3.5 A particle of mass m_e moves in a three-dimensional spherically symmetric well where $V = 0$, $r \leqslant a$ and $V = V_0$, $r > a$. Show that the energies of those states with quantum number $l = 0$ are determined by the condition $k\cot ka = -\kappa$ in the notation used in chapter 2. Show that there are no bound states of such a system unless $V_0 > \hbar^2\pi^2/8ma^2$. Would you expect this condition to be modified if states with $l \neq 0$ were also considered?

3.6 The (negative) binding energy of the ground state of the deuteron (neutron+proton) is 2.23 MeV. Assuming that the interaction potential is of the form described in problem 3.5 with $a = 2.0 \times 10^{-15}$ m, find the corresponding value of V_0. Do bound states of the deuteron other than the ground state exist?

Hints: Use the reduced mass (see footnote 4); $x = 1.82$ is a solution to the equation $x\cot x = -0.46$.

3.7 Verify that

$$\int_0^{2\pi}\int_0^{\pi} Y_{lm}^* Y_{l'm'}\sin\theta\, d\theta\, d\phi = 0 \qquad \text{unless } l = l' \text{ and } m = m'$$

for all values of l, l', m and m' up to and including those with l and/or l' equal to 2, where Y_{lm} is a spherical harmonic and Y_{lm}^* its complex conjugate.

3.8 Use the hydrogen atom wavefunctions and the probabilistic interpretation of the wavefunction to calculate (i) the most probable and (ii) the average value, of the distance between the electron and proton in a hydrogen atom in its 1*s* state.

Chapter 4

The basic postulates of quantum mechanics

In the previous chapters we have seen how solutions to the time-independent Schrödinger equation correspond to the allowed energy levels of a quantum system and how, in the hydrogen atom case in particular, the results of this procedure are in extremely good agreement with experiment. It would be possible to extend the process to a prediction of the energy levels of other atoms. We would find that the corresponding Schrödinger equations could no longer be solved exactly, but that approximations could be developed which, when combined with computational techniques, would lead to predicted energy levels that were once again in very good agreement with experiment. However, such a programme, most of which is beyond the scope of this book in any case, would be premature at this stage as we have not yet established a general procedure which will do more than predict the allowed energy levels of a particle moving in a potential and the probability that it is in the vicinity of a particular point in space. We do not yet know, for example, how to predict the momentum of an electron in a hydrogen atom; we do not even know if it has a definite value or if it can only be specified by a probability distribution, as is the case for electron position. We have as yet no way of predicting under what conditions an atom undergoes a transition from one state to another, emitting energy in the form of a quantum of electromagnetic radiation.

Answers to this and similar questions require a more general and more sophisticated approach to quantum mechanics. We shall develop the foundations of this in the present chapter and we shall see that it is based on five postulates. In developing and discussing the postulates, we shall build on what we have already done in previous chapters to make each postulate seem at least reasonable. It is important to remember that this is an *inductive* process; i.e. although the postulates can be made to appear reasonable by considering particular examples, they can never be rigorously derived in this way. We encountered a previous example of an inductive argument when we set up the Schrödinger equation on the basis of a consideration of classical waves and the experimental de Broglie relations; we emphasized then that this argument in no way represented a proof

of the Schrödinger equation and that our belief in its correctness lay in the fact that it was successful in predicting the results of experiments. In a similar way the correctness or otherwise of the basic postulates of quantum mechanics rests on the agreement between deductions from them and the results of experiments. To the present time quantum mechanics has withstood every experimental test; there are many fields of physics that are not yet completely understood, but there are no experimental results that contradict or falsify the fundamental principles of quantum mechanics. Despite this, the fundamental basis of the subject contains ideas that run very much against the intuition we have all developed in our interaction with the classical world. We have come across some of these already in our discussion of wave–particle duality in earlier chapters and we shall come across some more in this chapter and later. However, we shall postpone any detailed discussion of the foundations of the subject to the last chapter, where we shall attempt to introduce the reader to the main ideas underlying the vigorous philosophical debate about the conceptual basis of quantum mechanics that has gone on ever since the early days of the subject.

It might therefore have been expected that we should simply have stated the basic postulates at the beginning of this book, thereafter concentrating on deriving predictions of experimental results from them, and indeed such an approach is sometimes adopted. However, the ideas of quantum mechanics are rather abstract and most students understand them more easily when introduced to them gradually using inductive arguments based on a knowledge of the wave mechanics of atoms as described in the previous chapters.

4.1 The wavefunction

The experimental evidence for the wave properties of the electron and the success of the Schrödinger equation in predicting the energy levels of the hydrogen atom indicate that the physical properties of a particle moving under the influence of a potential can be obtained from the *wavefunction*. As we saw, this is a single-valued function of the coordinates of the particle and the time; it is finite, differentiable and its squared modulus must be integrable over all possible particle positions. It is important to remember that, although the measurable properties of a system are derived from the wavefunction, it is not itself a physical quantity. However, it is a fundamental principle of quantum mechanics that the wavefunction contains all the information it is possible to obtain about a physical system in a particular state. The first postulate then concerns the existence of the wavefunction.

Postulate 4.1 For every dynamical system there exists a wavefunction that is a continuous, square-integrable, single-valued function of the parameters of the system and of time, and from which all possible predictions about the physical properties of the system can be obtained.

This statement covers the case of a particle moving in a potential where the 'parameters of the system' are the particle coordinates, but it also refers to more general situations: for example, the parameters may be the coordinates of all the particles of a many-body system and may include internal variables such as 'spin'.

Notation

In the previous chapters we used the symbol Ψ to represent a general solution to the time-dependent Schrödinger equation and the symbol u (sometimes with a subscript) as the time-independent part of the wavefunction of a system in a state of given energy. We shall continue this notation in the present chapter, and in addition we shall use the symbol ψ to represent a general wavefunction, whose time dependence we are not explicitly considering, and the symbol ϕ (often with a subscript) when the system is in what we shall call an 'eigenstate'—that is when some dynamical quantity (not necessarily the energy) has a known value.

4.2 The dynamical variables

In this section we consider how dynamical variables (position, momentum, angular momentum, energy, etc.) are represented mathematically in quantum mechanics. In classical mechanics, such quantities are represented by algebraic variables: position coordinates x, y, z; momentum components p_x, p_y, p_z; energy E; etc. However, algebraic variables can take on any value whereas we have seen that quantum-mechanical properties are often confined to a discrete set of values (e.g. the energy levels of a hydrogen atom). Moreover, algebraic variables imply precise values, while the uncertainty principle indicates that simultaneous specification of the magnitudes of two dynamical quantities (e.g., position and momentum) is not always possible in quantum mechanics. We therefore look for a new way of representing dynamical quantities mathematically.

Let us consider again the time-independent Schrödinger equation in the case of a particle moving in a potential $V(\mathbf{r})$, which was given in equation (3.7) and which we now write in the form

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right] u_n = E_n u_n \tag{4.1}$$

or

$$\hat{H}u_n = E_n u_n \qquad \text{where } \hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V \tag{4.2}$$

The quantity in square brackets in (4.1), which is defined as $\hat{H}$ in (4.2), is an example of a mathematical *operator*. Such operators operate on functions (in this case the u_n) to produce new functions, and equation (4.2) therefore states that the energy levels E_n and corresponding wavefunctions u_n are such that, when the operator $\hat{H}$ operates on u_n, it produces a result equivalent to multiplying u_n

by the constant E_n. The quantities E_n and the functions u_n are known as the *eigenvalues* and *eigenfunctions* respectively of the operator $\hat{H}$, and we say that the energy of the quantum-mechanical system is represented by an operator $\hat{H}$ whose eigenvalues are equal to the allowed values of the energy of the system. For historical reasons $\hat{H}$ is often known as the *Hamiltonian operator*.

We now consider the significance of the eigenfunctions u_n. In the earlier chapters we interpreted u_n as representing the wavefunction of the system when it was in a state whose energy is E_n. This implies that if the energy of the system were to be measured when the wavefunction is u_n, we should certainly obtain the result E_n. As a second measurement of the energy performed immediately after the first would be reasonably expected to yield the same result, we conclude that the wavefunction of a quantum-mechanical system will be identical to the corresponding eigenfunction of the Hamiltonian operator *immediately after* a measurement of the energy of the system.

It is a reasonable extension of these arguments to say that other dynamical variables (position, momentum, etc.) should also be capable of representation by operators, that the eigenvalues of these operators should correspond to the possible results of experiments carried out to measure these quantities, and that the wavefunction of a quantum-mechanical system should be identical to the corresponding eigenfunction following the measurement represented by such an operator. Consider first the momentum operator: the first part of the Hamiltonian (4.2) corresponds to the kinetic energy of the particle, and classically this is related to the particle momentum by the expression

$$T = p^2/2m \tag{4.3}$$

If we assume that a similar relation holds in quantum mechanics we get

$$\frac{1}{2m}\hat{P}^2 = -\frac{\hbar^2}{2m}\nabla^2 \tag{4.4}$$

where $\hat{\mathbf{P}}$ is the momentum operator. An expression for $\hat{\mathbf{P}}$ which is consistent with this is

That is,

$$\left.\begin{aligned} \hat{\mathbf{P}} &= -i\hbar\nabla \\ \hat{P}_x &= -i\hbar\frac{\partial}{\partial x} \qquad \text{etc.} \end{aligned}\right\} \tag{4.5}$$

where $\hat{P}_x$ is the operator representing the x component of the momentum. The negative sign in the definition of $\hat{P}_x$ is chosen to ensure consistency with our earlier results as we shall soon see.

We now consider the eigenvalue equation of the operator representing the x component of momentum: if the eigenvalues and eigenfunctions are p and ϕ respectively, then

$$-i\hbar\frac{\partial}{\partial x}\phi = p\phi \tag{4.6}$$

that is,

$$\phi = A\exp(ikx) \tag{4.7}$$

where A is a constant and $k = p/\hbar$. But this is just the de Broglie relation connecting wave number and momentum (2.1); and equation (4.7) is therefore consistent with the discussion in earlier chapters. Moreover, this explains the choice of sign for the momentum operator (4.5).

We now note two points about the momentum eigenvalue equation. First, there is a solution to (4.6) for any value of p so the possible values of the momentum of a particle are not confined to a discrete set in the way that the allowed energy levels of a bound particle are. Second (apart from the case of a free particle when $V(r)$ is constant everywhere), the momentum eigenfunctions are different from the energy eigenfunctions. Thus if the energy of a bound particle is measured, the wavefunction immediately after the measurement will not be an eigenfunction of the momentum operator. It follows that the outcome of a measurement of the momentum of a bound particle when the energy is already known cannot be accurately predicted. We shall return to this point in more detail later when we discuss the fourth postulate.

We now consider the operator representing the position of a particle, confining our discussion to one dimension for the moment. The eigenfunction equation can be written formally as

$$\hat{X}\phi = x_0\phi \tag{4.8}$$

where x_0 and ϕ now represent the position eigenvalues and eigenfunctions respectively and we have to obtain a form of $\hat{X}$ which is consistent with what we already know. We remember that in the previous chapters the squared modulus of the wavefunction was interpreted as the probability density for the position of the particle. If, therefore, the eigenfunction ϕ is to represent the wavefunction immediately after a position measurement yielding the result x_0, it follows that ϕ must be very large inside a region of very narrow width Δ in the vicinity of x and zero elsewhere. In the limit as Δ goes to zero, the wavefunction has the form of a *Dirac delta function* and is written as $\delta(x - x_0)$. In this case

$$\hat{X}\delta(x - x_0) = x_0\delta(x - x_0) \tag{4.9}$$

which equation is satisfied if

$$\hat{X} \equiv x \tag{4.10}$$

as can be seen by substituting (4.10) into (4.9) and considering separately the point $x = x_0$ (where the expressions on the two sides of (4.9) are identical) and the region $x \neq x_0$ (where both sides are zero). That is, the quantum-mechanical operator representing the x coordinate of a particle is just the algebraic variable x. In three dimensions, it follows that the particle position operator is the vector $\mathbf{r}$ and we note that this is consistent with the fact that we wrote the energy operator in the form

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$$

where $V(\mathbf{r})$ is just the potential expressed as a function of the algebraic variable $\mathbf{r}$. We thus conclude that the position and momentum of a particle can be represented by the operators $\mathbf{r}$ and $-i\hbar\nabla$ respectively. We saw that the kinetic and potential energy operators have the same functional dependence on the operators representing position and momentum as do the corresponding quantities in classical mechanics and we assume that this relationship also holds for other dynamical quantities which can be expressed classically as functions of $\mathbf{r}$ and $\mathbf{p}$. We write the eigenfunction equation in the case of a general operator $\hat{Q}$ as

$$\hat{Q}\phi_n = q_n\phi_n \tag{4.11}$$

Before expressing all this in the form of formal postulates, we shall establish one general property of operators used to represent dynamical variables. We have interpreted the eigenvalues of the operators discussed so far as representing the possible results of experiments carried out to measure the values of the corresponding physical quantities. If this interpretation is to be correct, it is clear that these eigenvalues must be real numbers even though the eigenfunctions, or the operators themselves, may be imaginary (like the momentum operator) or complex. One class of mathematical operators that always have real eigenvalues consists of the *Hermitian* operators. Hermitian operators are defined such that if $f(\mathbf{r})$ and $g(\mathbf{r})$ are any well-behaved functions of $\mathbf{r}$ which vanish at infinity then the operator $\hat{Q}$ is Hermitian if and only if

$$\int f\hat{Q}g\,d\tau = \int g\hat{Q}^* f\,d\tau \tag{4.12}$$

where $\hat{Q}^*$ is the complex conjugate of $\hat{Q}$ and the integrals are over all values of $\mathbf{r}$. The complex conjugate of an operator is defined so that, for any function f, $\hat{Q}^* f^*$ must equal $(\hat{Q}f)^*$; in nearly all cases, this definition is equivalent to replacing i with $-i$ wherever it appears in $\hat{Q}$.

Readers familiar with Hermitian matrices may like to note that (4.12) is equivalent to

$$Q_{12} = Q_{21}^* \tag{4.13}$$

where

$$Q_{12} = \int f_1^*\hat{Q}f_2\,d\tau \tag{4.14}$$

(4.13) follows directly from (4.12) if we replace f and g by f_1^* and f_2 respectively.

We now show that the eigenvalues of a Hermitian operator are real.

Let

$$\hat{Q}\phi_n = q_n\phi_n$$

then

$$\hat{Q}^*\phi_n^* = q_n^*\phi_n^*$$

Hence

$$\int \phi_n^* \hat{Q} \phi_n \, d\tau = q_n \int \phi_n^* \phi_n \, d\tau \tag{4.15}$$

and

$$\int \phi_n \hat{Q}^* \phi_n^* \, d\tau = q_n^* \int \phi_n \phi_n^* \, d\tau \tag{4.16}$$

But if $\hat{Q}$ is Hermitian it follows from (4.12) (identifying f with ϕ_n and g with ϕ_n^*) that the left-hand sides of (4.15) and (4.16) must be equal. Equating the right-hand sides of these equations we get

$$q_n \int |\phi_n|^2 \, d\tau = q_n^* \int |\phi_n|^2 \, d\tau \tag{4.17}$$

and hence

$$q_n = q_n^* \tag{4.18}$$

so that q_n is real as required.

The converse of this theorem (that all operators with real eigenvalues are Hermitian) is not true: the Hermitian property is a stronger condition on the operator than the reality of the eigenvalues. However, as it has been found possible to represent all physical quantities by Hermitian operators, and as these have a number of useful properties, some of which will be discussed later, it is convenient to impose this Hermitian condition.

We can readily show that the operators representing position and momentum in one dimension are Hermitian: in the case of x this follows trivially from substitution into the one-dimensional equivalent of (4.12) while in the case of $\hat{P}_x$ we have

$$\int_{-\infty}^{\infty} f \hat{P}_x g \, dx = -i\hbar \int_{-\infty}^{\infty} f \frac{\partial g}{\partial x} \, dx$$

Integrating by parts, the right-hand side becomes

$$-i\hbar \left\{ \Big[fg \Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} g \frac{\partial f}{\partial x} \, dx \right\}$$

The first term is zero because f and g vanish at infinity so we have

$$\begin{aligned} \int_{-\infty}^{\infty} f \hat{P}_x g \, dx &= i\hbar \int_{-\infty}^{\infty} g \frac{\partial f}{\partial x} \, dx \\ &= \int_{-\infty}^{\infty} g \hat{P}_x^* f \, dx \end{aligned}$$

as required. We leave it as an exercise for the reader to extend this argument to three dimensions and to confirm that other operators considered so far are Hermitian.

We can now summarize the contents of these paragraphs in two postulates.

Postulate 4.2 Every dynamical variable may be represented by a Hermitian operator whose eigenvalues represent the possible results of carrying out a measurement of the value of the dynamical variable. Immediately after such a measurement, the wavefunction of the system will be identical to the eigenfunction corresponding to the eigenvalue obtained as a result of the measurement.

Postulate 4.3 The operators representing the position and momentum of a particle are $\mathbf{r}$ and $-i\hbar\nabla$ respectively. Operators representing other dynamical quantities bear the same functional relation to these as do the corresponding classical quantities to the classical position and momentum variables.

Orthonormality

One important property of the eigenfunctions of a Hermitian operator is known as *orthonormality* and is expressed by the following relation:

$$\int \phi_n^* \phi_m \, d\tau = \delta_{nm} \tag{4.19}$$

where ϕ_n and ϕ_m are eigenfunctions of some Hermitian operator and δ_{nm} is the *Kronecker delta* defined by $\delta_{nm} = 0,\ n \neq m;\ \delta_{nm} = 1,\ n = m$. To prove this consider a Hermitian operator $\hat{Q}$ whose eigenvalues are q_n and whose eigenfunctions are ϕ_n. That is,

$$\hat{Q}\phi_n = q_n \phi_n \tag{4.20}$$

Thus

$$\int \phi_n^* \hat{Q} \phi_m \, d\tau = q_m \int \phi_n^* \phi_m \, d\tau \tag{4.21}$$

(because q_m is a constant which can be taken outside the integral) and similarly (remembering that q_n is real)

$$\int \phi_m \hat{Q}^* \phi_n^* \, d\tau = q_n \int \phi_m \phi_n^* \, d\tau \tag{4.22}$$

But it follows from the definition of the Hermitian operator (4.12) that the left-hand sides of (4.21) and (4.22) are equal. We thus have, equating the right-hand sides,

$$\int \phi_n^* \phi_m \, d\tau = 0 \qquad \text{or} \qquad q_m = q_n \tag{4.23}$$

The second alternative implies that $m = n$ or that the eigenfunctions ϕ_m and ϕ_n correspond to the same eigenvalue—i.e. that they are degenerate. We postpone a discussion of degeneracy until later in this chapter and conclude for the moment that

$$\int \phi_n^* \phi_m \, d\tau = 0 \qquad m \neq n \tag{4.24}$$

Turning now to the case $m = n$, it is clear from the eigenvalue equation (4.20) that if ϕ_n is an eigenfunction of $\hat{Q}$ then so is $K\phi_n$ where K is any constant. As $|\phi_n|^2 \, d\tau$ represents the probability of finding the particle in the volume element $d\tau$ and as the total probability of finding the particle somewhere in space is unity, we choose the scaling constant so that

$$\int |\phi_n|^2 \, d\tau = 1 \tag{4.25}$$

Equations (4.24) and (4.25) are clearly equivalent to (4.19) so we have proved orthonormality in the case of non-degenerate eigenfunctions. We shall see later that, although degenerate eigenfunctions are not necessarily orthogonal, an orthonormal set of wavefunctions can always be chosen in the degenerate case also.

4.3 Probability distributions

So far we have postulated that physical quantities can be represented by Hermitian operators whose eigenvalues represent the possible results of experimental measurements and whose corresponding eigenfunctions represent the wavefunction of the system immediately after the measurement. Clearly if, for example, the wavefunction of a system is also identical to one of the energy eigenfunctions u_n immediately *before* a measurement of the energy of the system, then the result E_n will definitely be obtained; however, we have not yet postulated what the outcome of the experiment will be if the wavefunction before the measurement is *not* one of the u_n. The first point to be made is that, in this case, quantum mechanics does not make a precise prediction about the result of the energy measurement. This is not to say that a precise *measurement* of the energy is impossible (in fact postulate 4.2 clearly states that only a result precisely equal to one of the energy eigenvalues *is* possible) but that the result of such a measurement is *unpredictable*. Clearly, the measurement of energy is just one example and similar reasoning can be applied to the measurement of other physical properties. The inability of quantum mechanics to predict the actual outcome of many individual physical events is a fundamental feature of the theory. However, the relative probabilities of the different possible outcomes can always be predicted, and we shall now proceed to develop a postulate concerning these probabilities and how they can be derived from a knowledge of the wavefunction and the operators representing the dynamical variables.

We have already seen (cf. Section 2.1) how knowledge of the wavefunction can be used to predict the relative probabilities of the possible outcomes of a measurement of the position of a particle. We said there that if the wavefunction of the system before the position measurement is carried out is $\psi(\mathbf{r})$, then the probability that the position measurement will yield a result within the element of volume $d\tau$ around $\mathbf{r}$ is $|\psi(\mathbf{r})|^2 \, d\tau$. The outcome of a position measurement

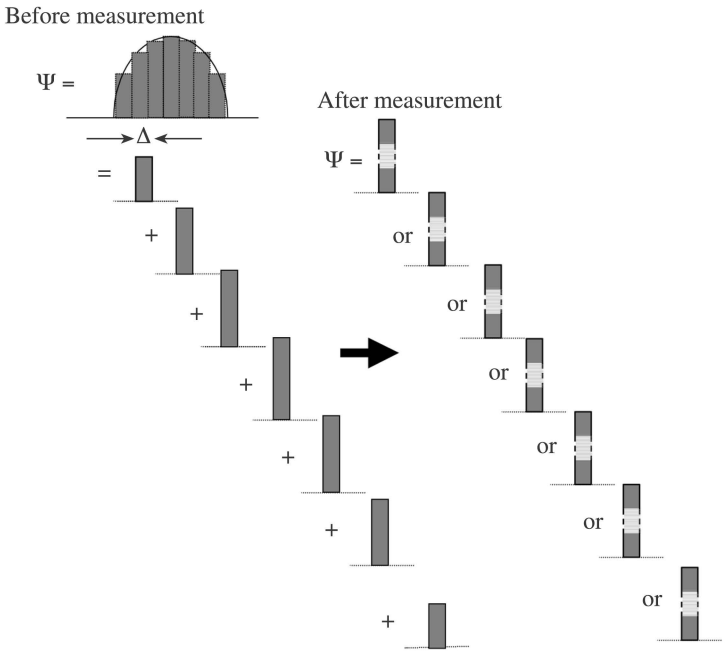


Figure 4.1. An example of quantum measurement theory. The left-hand part of the diagram shows how the wavefunction before the measurement can be expressed as a sum of parts, each of which corresponds to one of the seven possible outcomes. After the measurement the wavefunction 'collapses' at random into one of the functions shown on the right. (NB: The vertical scale of the right-hand peaks is reduced.)

is exactly predictable only if the wavefunction is zero everywhere except at a particular point where it is very large. We saw in the previous section that this 'Dirac-delta' form is an eigenfunction of the position operator **r**.

Dirac deltas give rise to technical problems that are best avoided at this stage and in any case no practical measurement can determine the particle position with infinite accuracy. Instead, we consider an experiment where the particle is found in one of a series of detectors arranged along the x axis, each one having a small, but finite, width Δ as in figure 4.1. *After* the measurement, the particle is found in one of the detectors, so its wavefunction must be zero everywhere except within a peak of height h (*not* Planck's constant) and width Δ. The total probability of finding the particle somewhere must be unity, so it follows that

$$h^2\Delta = 1 \qquad \text{and} \qquad h = \Delta^{-1/2} \tag{4.26}$$

We denote the nth such wavefunction as ϕ_n

Suppose now that the wavefunction *before* the measurement is $\psi(x)$. We imagine $\psi(x)$ divided into a series of strips of width Δ corresponding to the

particle detectors (see figure 4.1 again) so we can approximate $\psi(x)$ by $\psi(x_n)$ inside the nth strip. This approximation becomes exact in the limit of small Δ. Remembering that ϕ_n represents a peak of height h covering this strip, we can write

$$\begin{aligned}\psi(x) &= \sum_n \psi(x_n)\phi_n/h \\ &= \sum_n a_n\phi_n \end{aligned} \tag{4.27}$$

where

$$a_n = \psi(x_n)/h = \Delta^{1/2}\psi(x_n) \tag{4.28}$$

We already know the probability of finding a particle in the element dx at x is $|\psi(x)|^2$, so the probability of finding the particle within the strip at x_n is just $|\psi(x_n)|^2\Delta$, which equals $|a_n|^2$.

We now generalize this to a measurement represented by the operator $\hat{Q}$ (where $\hat{Q}\phi_n = q_n\phi_n$) on a system whose wavefunction is known to be ψ immediately before the measurement. We postulate that the above procedure can be applied quite generally so that

$$\psi = \sum_n a_n\phi_n \tag{4.29}$$

(where the summation is over all the eigenfunctions of the operator) and the probability of obtaining the result q_n, when the quantity represented by $\hat{Q}$ is measured, is $|a_n|^2$. The expression (4.29) relies on a mathematical property of the eigenfunctions of a Hermitian operator known as *completeness*. This states that any well-behaved function, such as the wavefunction ψ, can be expressed as a linear combination of the eigenfunctions ϕ_n which are then said to form a *complete set*.

We now re-state this as the fourth postulate.

Postulate 4.4 When a measurement of a dynamic variable represented by the Hermitian operator $\hat{Q}$ is carried out on a system whose wavefunction is ψ, then the probability of the result being equal to a particular eigenvalue q_m will be $|a_m|^2$, where $\psi = \sum_n a_n\phi_n$ and the ϕ_n are the eigenfunctions corresponding to the eigenvalues q_n.

We now show that in the case where both the wavefunction ψ and the eigenfunctions ϕ_n are known, we can use orthonormality to obtain an expression

for a_n:

$$\begin{aligned}\int \phi_n^* \psi \, d\tau &= \int \phi_n^* \sum_m a_m \phi_m \, d\tau \\ &= \sum_m a_m \int \phi_n^* \phi_m \, d\tau \\ &= \sum_m a_m \delta_{nm} = a_n \end{aligned} \tag{4.30}$$

So that

$$a_n = \int \phi_n^* \psi \, d\tau \tag{4.31}$$

The right-hand-side of (4.31) is sometimes called the *overlap integral* connecting ψ and ϕ_n. Essentially it is a measure of the extent to which the two functions are similar, so this makes some sense of the fact that its square is the probability of obtaining the corresponding result of the measurement.

We developed our argument by considering the measurements of position, so let us see what it now gives us for momentum measurements. Suppose we perform an experiment to measure the momentum of a particle as having one of the values p_n, where these are a set of closely spaced bands—similar to those considered in the earlier position measurement. Applying postulate 4.4, we say that the probability of obtaining the result $p_n (= \hbar k_n)$ is $|a_n|^2$ where, using the momentum eigenfunctions (4.7),

$$\psi(x) = \sum_n a_n A \exp i k_n x \tag{4.32}$$

We see that in this case, completeness leads to an expression that is just the Fourier expansion of the function $\psi(x)$. Applying (4.31) to the momentum case, we get

$$a_n = \int \psi(x) A \exp(-i k_n x) dx \tag{4.33}$$

which is the standard expression for the Fourier amplitude.

Continuous eigenvalues

The previous arguments have been developed on the assumption that the eigenvalues of the operator in question have discrete values that can be indexed by an integer such as n. This is true for the energy levels of a particle bound to a potential and (as we shall see later) for the angular momentum eigenvalues, but other quantities, notably position and momentum, have a continuous range of eigenvalues. There are several ways by which these cases can be brought within our formalism. One way is to 'force' the problem to be discrete. We did this earlier in the case of position by considering narrow ranges of position instead

of infinitesimal points. Another way of doing this is to imagine the system to be contained inside a large box, so that the eigenvalues are closely spaced, but not continuous. Outside the box, the wavefunction may be assumed to be zero or we may imagine that its form inside the box is repeated periodically through all space. The latter device ensures that the wavefunction can be expanded as a Fourier series, rather than a transform and this is effectively what we did in (4.32). An alternative is to adapt our formalism to the continuous case by making use of the properties of Dirac-delta functions. Writing the general eigenvalue equation (4.11) as

$$\hat{Q}\phi(k, x) = q(k)\phi(k, x) \tag{4.34}$$

where the summation index, n is replaced by the continuous variable k, the equivalent of (4.29) is

$$\psi(x) = \int a(k)\phi(k, x)dk \tag{4.35}$$

Instead of $|a_n|^2$ as the probability of getting the result q_n we have $|a(k)|^2 dk$ as the probability of getting a result inside the range dk in the vicinity of k. The orthonormality condition (4.19) then becomes

$$\int \phi^*(k, x)\phi(k', x)dx = \delta(k - k') \tag{4.36}$$

As an example, we show how momentum eigenfunctions can be treated in this way, restricting ourselves to one dimension. We put

$$\phi(k, x) = (2\pi)^{-1/2} \exp(ikx)$$

so that

$$\psi(x) = (2\pi)^{-1/2} \int a(k) \exp(ikx)dk$$

and

$$a(k) = (2\pi)^{-1/2} \int \psi(x) \exp(ikx)dx \tag{4.37}$$

$a(k)$ is therefore the Fourier transform of $\psi(x)$ and can be thought of as a 'wavefunction in k space'.

It should be noted that k may represent a set of variables, rather than a single variable. For example, representation of the momentum of a three-dimensional system requires a vector $\mathbf{k}$ with three components. Similarly, in many discrete systems the index n used before represents a set of several integer indices.

Summary

We have now set out the core ideas of quantum measurement theory and the reader would be wise to pause at this point and ensure that they have been grasped. Putting it all a little less formally:

1. A quantum system has a wavefunction associated with it.

2. When a measurement is made, the result is one of the eigenvalues of the operator associated with the measurement.

3. As a result of the measurement the wavefunction 'collapses' into the corresponding eigenfunction.

4. The probability of a particular outcome equals the squared modulus of the overlap between the wavefunctions before and after the measurement.

Look again at figure 4.1 and make sure you understand how this applies to the case of a position measurement.

Expectation values

We can use this formalism to predict the average value that will be obtained from a large number of measurements of the same quantity. We assume that the wavefunctions before each experiment are identical. An example would be a stream of particles whose positions were all measured with an apparatus such as that in figure 4.1. This average is known as the *expectation value* and is obtained in the case of a wavefunction ψ and an operator $\hat{Q}$ as follows. Consider the expression

$$\int \psi^* \hat{Q} \psi \, d\tau$$

Using the expansion of ψ in terms of the eigenfunctions ϕ_n of $\hat{Q}$ (4.29) we can write this as

$$\begin{aligned} \int \psi^* \hat{Q} \psi \, d\tau &= \int \left(\sum_m a_m^* \phi_m^* \right) \hat{Q} \left(\sum_n a_n \phi_n \right) d\tau \\ &= \sum_{m,n} a_m^* a_n q_n \int \phi_m^* \phi_n \, d\tau \\ &= \sum_{m,n} a_m^* a_n q_n \delta_{mn} \\ &= \sum_n |a_n|^2 q_n \end{aligned} \tag{4.38}$$

But postulate 4.4 states that $|a_n|^2$ is just the probability that the value q_n be obtained in the measurement, so the left-hand side of (4.38) is clearly the expectation value we are looking for. This is often written as $\langle \hat{Q} \rangle$ so we have

$$\langle \hat{Q} \rangle = \int \psi^* \hat{Q} \psi \, d\tau \tag{4.39}$$

Thus, if we know the wavefunction of the system and the operator representing the dynamical variable being measured, we can calculate the expectation value.

4.4 Commutation relations

Consider the following expression where ψ is any wavefunction, and $\hat{X}$ and $\hat{P}_x$ are the operators representing the x coordinate and x component of momentum of a particle:

$$\begin{aligned}(\hat{P}_x\hat{X} - \hat{X}\hat{P}_x)\psi &= -i\hbar\frac{\partial}{\partial x}(x\psi) - x\left(-i\hbar\frac{\partial\psi}{\partial x}\right) \\ &= -i\hbar x\frac{\partial\psi}{\partial x} - i\hbar\psi + i\hbar x\frac{\partial\psi}{\partial x} \\ &= -i\hbar\psi \end{aligned} \tag{4.40}$$

We note two important points. First, the effect on ψ of the product of two quantum-mechanical operators is, in general, dependent on the order in which the operators are applied: that is, unlike algebraic variables, quantum-mechanical operators do not in general commute. Second, the result (4.40) is completely independent of the particular form of ψ so we can write

$$[\hat{P}_x, \hat{X}] \equiv \hat{P}_x\hat{X} - \hat{X}\hat{P}_x = -i\hbar \tag{4.41}$$

where $[\hat{P}_x, \hat{X}]$ is defined by (4.41) and is known as the *commutator* or *commutator bracket* of $\hat{P}_x$ and $\hat{X}$. Similar arguments to these can be used to show that

$$[\hat{P}_y, \hat{Y}] = [\hat{P}_z, \hat{Z}] = -i\hbar$$

and that commutators of the form $[\hat{X}, \hat{Y}]$, $[\hat{P}_x, \hat{Y}]$, $[\hat{P}_x, \hat{P}_y]$ etc., are all equal to zero. All the commutators relating to the different components of position and momentum can therefore be collected together in the following statements

$$\left.\begin{aligned}[\hat{X}_i, \hat{X}_j] = [\hat{P}_i, \hat{P}_j] &= 0 \\ [\hat{P}_i, \hat{X}_j] = -[\hat{X}_j, \hat{P}_i] &= -i\hbar\delta_{ij}\end{aligned}\right\} \tag{4.42}$$

where we are now using the notation $\hat{X}_1 \equiv \hat{X}$; $\hat{X}_2 \equiv \hat{Y}$; etc.

Expressions for the commutation relations between other pairs of operators representing dynamical variables can be obtained in a similar manner, although not every commutator is equal to a constant as was the case for $[\hat{P}_x, \hat{X}]$, etc. For example, the commutator bracket of the x coordinate of a particle and its kinetic energy $[\hat{X}, \hat{T}]$ can be shown to be the operator $i\hbar\hat{P}_x/m$. Commutation relations are generally of great importance in quantum mechanics. In fact it is possible to use the expressions (4.42) to define the operators representing position and momentum rather than making the explicit identifications, $\hat{X} \equiv x$ and $\hat{P}_x = -i\hbar\partial/\partial x$, etc. Clearly other sets of operators could be chosen which would satisfy these commutation relations; for example, the momentum operator could be put equal to the algebraic variable p with the one-dimensional position operator being $i\hbar\partial/\partial p$, but it can be shown that the physical predictions of quantum mechanics

are independent of the choice of operators provided the commutation relations (4.42) are satisfied. It is possible to develop much of the formal theory of quantum mechanics using the commutation relations without postulating particular forms for the operators and this 'representation-free' approach is employed in many advanced texts. However, although we shall prove several important relations without using explicit expressions for the operators, we shall use the 'position representation' set up in postulate 4.3 whenever this makes the argument clearer or more readily applicable to particular examples.

We shall now use the concept of the commutator bracket to discuss the ideas of compatible measurements and the uncertainty principle.

Compatibility

Two physical observables are said to be compatible if the operators representing them have a common set of eigenfunctions. This means that if one quantity is measured, the resulting wavefunction of the system will be one of the common eigenfunctions; a subsequent measurement of the other quantity will then have a completely predictable result and will leave the wavefunction unchanged. (As will be seen later, this statement has to be modified in the degenerate case.) We shall now show that the operators representing compatible measurements commute. Let the operators be $\hat{Q}$ and $\hat{R}$ with respective eigenvalues q_n and r_n and common eigenfunctions ϕ_n. Then, if ψ is any physical wavefunction which can therefore be written in the form (4.29)

$$\psi = \sum_n a_n \phi_n$$

we have

$$\begin{aligned}
[\hat{Q}, \hat{R}]\psi &= \sum_n a_n (\hat{Q}\hat{R}\phi_n - \hat{R}\hat{Q}\phi_n) \\
&= \sum_n a_n (\hat{Q} r_n \phi_n - \hat{R} q_n \phi_n) \\
&= \sum_n a_n (r_n q_n \phi_n - q_n r_n \phi_n) \\
&= 0
\end{aligned}$$

Thus $[\hat{Q}, \hat{R}] = 0$ if $\hat{Q}$ and $\hat{R}$ represent compatible observables. The converse can also be proved in the non-degenerate case (the degenerate case is discussed later in this chapter) as we now show. Given that the operators $\hat{Q}$ and $\hat{R}$ commute, let ϕ_n be an eigenfunction of $\hat{Q}$ so that

$$\hat{Q}\hat{R}\phi_n = \hat{R}\hat{Q}\phi_n = \hat{R} q_n \phi_n = q_n \hat{R}\phi_n \tag{4.43}$$

It follows directly from (4.43) that $(\hat{R}\phi_n)$ is also an eigenfunction of $\hat{Q}$ with eigenvalue q_n. In the absence of degeneracy therefore $(\hat{R}\phi_n)$ can differ from ϕ_n

only by a multiplicative constant. If we call this constant r_n we have

$$\hat{R}\phi_n = r_n\phi_n$$

and ϕ_n must therefore be an eigenfunction of both $\hat{Q}$ and $\hat{R}$. Thus, in order to test whether two quantities are compatible, it is sufficient to calculate the commutator of the operators representing them and check whether or not it is equal to zero. For example, it follows from this and from the commutation relations (4.42) that measurements of the three positional coordinates can be made compatibly, as can measurements of the three components of momentum. Moreover, a momentum component can be measured compatibly with the measurement of either of the other two position coordinates, but not with the position coordinate in the same direction. In the case of a free particle, the energy operator $(\hat{P}^2/2m)$ commutes with the momentum operator $(\hat{P})$ and these two quantities can therefore be measured compatibly: the common eigenfunctions in this case are plane waves of the form (4.7).

4.5 The uncertainty principle

One of the most well-known ideas in quantum mechanics is Heisenberg's uncertainty principle, which we discussed briefly in chapter 1. Arguably, it is also the most misunderstood idea in the field and we shall try to clear up some of this confusion later in this section. However, before doing so we shall use the formalism developed so far to obtain a general form of the uncertainty principle applicable to any pair of measurements.

We first need to establish some properties of pairs Hermitian operators. We will show that if $\hat{Q}$ and $\hat{R}$ are Hermitian operators then, although the product $\hat{Q}\hat{R}$ need not be Hermitian, the expressions $(\hat{Q}\hat{R} + \hat{R}\hat{Q})$ and $i(\hat{Q}\hat{R} - \hat{R}\hat{Q})$ are Hermitian. We have, using (4.12) and the fact that $\hat{Q}$ and $\hat{R}$ are Hermitian,

$$\begin{aligned}\int f\hat{Q}\hat{R}g\,d\tau &= \int (\hat{R}g)\hat{Q}^* f\,d\tau \\ &= \int (\hat{Q}^* f)\hat{R}g\,d\tau \\ &= \int g\hat{R}^*\hat{Q}^* f\,d\tau \end{aligned} \tag{4.44}$$

Similarly

$$\int f\hat{R}\hat{Q}g\,d\tau = \int g\hat{Q}^*\hat{R}^* f\,d\tau \tag{4.45}$$

Thus, adding (4.44) and (4.45)

$$\int f(\hat{Q}\hat{R} + \hat{R}\hat{Q})g\,d\tau = \int g(\hat{Q}\hat{R} + \hat{R}\hat{Q})^* f\,d\tau \tag{4.46}$$

while, subtracting (4.44) and (4.45) and multiplying by i,

$$\int f[i(\hat{Q}\hat{R} - \hat{R}\hat{Q}]g\, d\tau = \int g[i(\hat{Q}\hat{R} - \hat{R}\hat{Q})]^* f\, d\tau \tag{4.47}$$

which proves these statements. We note that it is an obvious corollary of this that if $\hat{Q}$ is Hermitian so is $\hat{Q}^2$. Thus, for example, the Hermitian property of the kinetic energy operator, $\hat{T} = \hat{P}^2/2m$, follows directly from the fact that $\hat{P}$ is Hermitian.

Consider now a series of measurements of the quantity represented by the operator $\hat{Q}$ on a system whose wavefunction is ψ before each measurement. The average result is equal to the expectation value $\langle\hat{Q}\rangle$ which was shown earlier (4.39) to be given by

$$\langle Q\rangle = \int \psi^* \hat{Q}\psi\, d\tau$$

We can also estimate the average amount by which the result of such a measurement would be expected to deviate from this expectation value: the operator representing this 'uncertainty' is clearly $(\hat{Q} - \langle\hat{Q}\rangle)$, so if the root-mean-square deviation from the mean is Δq, we have

$$\begin{aligned}\Delta q^2 &= \int \psi^*(\hat{Q} - \langle\hat{Q}\rangle)^2\psi\, d\tau \\ &= \int \psi^* \hat{Q}'\hat{Q}'\psi\, d\tau\end{aligned}$$

where $\hat{Q}'$ is defined as $(\hat{Q} - \langle\hat{Q}\rangle)$. Clearly if $\hat{Q}$ is Hermitian, so is $\hat{Q}'$ and therefore

$$\begin{aligned}\Delta q^2 &= \int (\hat{Q}'\psi)(\hat{Q}'^*\psi^*) \\ &= \int |\hat{Q}'\psi|^2\, d\tau\end{aligned} \tag{4.48}$$

In the same way, if we had carried out the measurement represented by the operator $\hat{R}$ on the same system, the expectation value would have been $\langle\hat{R}\rangle$ and the root-mean-square deviation Δr where

$$\Delta r^2 = \int |\hat{R}'\psi|^2\, d\tau$$

and

$$\hat{R}' = \hat{R} - \langle\hat{R}\rangle$$

Now consider the product

$$\begin{aligned}\Delta q^2 \Delta r^2 &= \int |\hat{Q}'\psi|^2\, d\tau \int |R'\psi|^2\, d\tau \\ &\geqslant \left| \int (\hat{Q}'^*\psi^*)(\hat{R}'\psi)\, d\tau \right|^2\end{aligned} \tag{4.49}$$

where the last step was obtained using a mathematical relationship known as Schwarz's inequality which states that

$$\int |f|^2 \, d\tau \int |g|^2 \, d\tau \geqslant \left| \int f^* g \, d\tau \right|^2$$

where f and g are any integrable functions of $\mathbf{r}$.[1] Now, again using the Hermitian property of $\hat{Q}'$,

$$\begin{aligned} \int (\hat{Q}'^* \psi^*)(\hat{R}'\psi) \, d\tau &= \int \psi^* \hat{Q}' \hat{R}' \psi \, d\tau \\ &= \tfrac{1}{2} \int \psi^* (\hat{Q}'\hat{R}' - \hat{R}'\hat{Q}') \psi \, d\tau \\ &\quad + \tfrac{1}{2} \int \psi^* (\hat{Q}'\hat{R}' + \hat{R}'\hat{Q}') \psi \, d\tau \end{aligned} \tag{4.50}$$

It was previously shown that if $\hat{Q}'$ and $\hat{R}'$ are Hermitian operators so are $(\hat{Q}'\hat{R}' + \hat{R}'\hat{Q}')$ and $i(\hat{Q}'\hat{R}' - \hat{R}'\hat{Q}')$ so the expectation values of the latter two operators must be real numbers. It follows that the first term on the right-hand side of (4.50) is purely imaginary and the second term is purely real, so the right-hand side of (4.49) can be expressed as the sum of squares of the two terms on the right-hand side of (4.50) and we get

$$\begin{aligned} \Delta q^2 \Delta r^2 &\geqslant \frac{1}{4} \left(\left| \int \psi^* [\hat{Q}', \hat{R}'] \psi \, d\tau \right|^2 + \left| \int \psi^* (\hat{Q}'\hat{R}' + \hat{R}'\hat{Q}') \psi \, d\tau \right|^2 \right) \\ &\geqslant \frac{1}{4} \left| \int \psi^* [\hat{Q}', \hat{R}'] \psi \, d\tau \right|^2 \end{aligned}$$

Now it follows immediately from the definitions of $\hat{Q}'$ and $\hat{R}'$ that $[\hat{Q}', \hat{R}'] = [\hat{Q}, \hat{R}]$ so we have

$$\Delta q \Delta r \geqslant \frac{1}{2} \left| \int \psi^* [\hat{Q}, \hat{R}] \psi \, d\tau \right|$$

[1] To prove Schwarz's inequality, consider the expression

$$\int \left| f \left(\int |g|^2 \, d\tau \right) - g \left(\int f g^* \, d\tau \right) \right|^2 d\tau$$

This expression must be greater than or equal to zero because the integrand is nowhere negative by definition. We can rewrite the integrand as a product of a function and its complex conjugate and get

$$\int \left[f^* \left(\int |g|^2 \, d\tau \right) - g^* \left(\int f^* g \, d\tau \right) \right] \left[f \left(\int |g|^2 \, d\tau \right) - g \left(\int f g^* \, d\tau \right) \right] d\tau \geqslant 0$$

When we multiply out the product of the two square brackets, two terms are equal and opposite and the others have a common factor, $\int |g|^2 \, d\tau$ that can be cancelled out. We then get

$$\int |f|^2 \, d\tau \int |g|^2 \, d\tau - \left| \int f^* g \, d\tau \right|^2 \geqslant 0$$

which is Schwarz's inequality.

That is

$$\Delta q \Delta r \geqslant \tfrac{1}{2}|\langle[\hat{Q}, \hat{R}]\rangle| \tag{4.51}$$

so, if we know the commutator of two operators, we can calculate the minimum value of the product of the uncertainties associated with series of measurements represented by each of them, on a system whose wavefunction is ψ before each measurement. This is known as the *generalized uncertainty principle*. In the case where $\hat{Q}$ and $\hat{R}$ are the operators representing a position coordinate and a corresponding component of momentum, we have $[\hat{X}, \hat{P}_x] = i\hbar$ and therefore

$$\Delta x \Delta p_x \geqslant \tfrac{1}{2}\hbar \tag{4.52}$$

which is the Heisenberg uncertainty principle referred to in chapter 1. We note that in this particular case the uncertainty product is independent of the wavefunction of the system.

It is tempting to deduce from (4.51) that, if two operators commute, the product of their uncertainties must be zero. However, this is not, in general, true because the uncertainties depend on the form of the initial wavefunction. Thus, for example, predictions of the values of the x and y coordinates of a particle initially in the ground state of the hydrogen atom both have uncertainties even though their operators commute. Nevertheless, we saw earlier that commuting operators represent compatible measurements, so in the particular case where the initial state is an eigenfunction of one of them, both uncertainties will be zero.

Thus we have shown that the uncertainty principle follows directly from the fundamental postulates of quantum mechanics. However, it is important to understand the nature of the 'uncertainties' Δq and Δr. First, these are *not* experimental errors associated with any particular measurement, but refer to the root-mean-square deviations or 'average spread' of a set of repeated measurements. This is the most we can predict about the outcome of the experiments, so we can look on the uncertainties as 'errors in *prediction*'. Second, the measurements represented by $\hat{Q}$ and $\hat{R}$ are obtained in separate sets of experiments, starting with the same wavefunction in each case. There is nothing in this to justify the popular notion that uncertainty arises because measuring one quantity 'disturbs' the system so as to introduce errors into the measurement of the other. These points should become clearer in the following example.

Example 4.1 Single-slit diffraction Consider a beam of particles of definite momentum p travelling parallel to the z axis (i.e. the wavefunction is a plane wave whose wavevector has components 0, 0, $p/\hbar$) towards a slit of width a in the x direction where the origin of coordinates is the centre of the slit (see figure 4.2). We consider how we can measure the x coordinate of the particle position and the x component of momentum after the particles pass through the slit. That is, the wavefunction ψ in this treatment corresponds to the wavefunction of a particle emerging from the slit. We first consider a set of experiments carried out to measure x, by placing a photographic film immediately behind the slit and

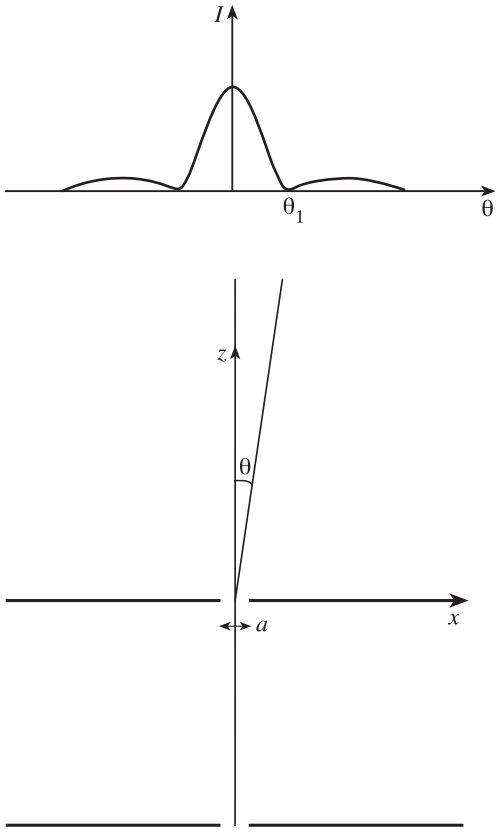


Figure 4.2. Single-slit diffraction. The lower slit ensures that particles reaching the upper slit are travelling parallel to the z axis. The graph at the top shows the intensity of the diffraction pattern as a function of θ in the case where $\lambda = a/3$.

allowing a large number of particles to pass through. If the beam is uniform across the slit, any particle has equal probability of passing through any part of the slit and it follows that the expectation value of its x coordinate will be zero with a standard deviation Δx of about $a/4$. (A direct application of (4.48) shows that its value is actually $a/2\sqrt{3}$.)

We measure the momentum uncertainty by a separate set of experiments in which we allow the particles to pass well beyond the slit before detecting them on the screen. Here, the wavefunction has the form of the well-known single-slit diffraction pattern, the first minimum of which occurs at an angle θ_1 where $\sin\theta_1 = \lambda/a = 2\pi\hbar/ap$. Thus most of the particles passing through the slit travel in a direction within the central maximum of the diffraction pattern, so we can predict that measurements of the x component of the momentum of the

particles will have an expectation value of zero with a spread Δp_x approximately equal to $p \sin\theta_1$. Hence

$$\Delta x \, \Delta p_x \simeq \frac{a}{2} p 2\pi \frac{\hbar}{ap} = \pi\hbar > \frac{1}{2}\hbar \tag{4.53}$$

in agreement with the uncertainty principle.

We should note a number of points about this example. First, when we record the diffraction pattern we can measure the x component of the particle momentum as precisely as we like by noting the point at which it blackens the film (any error in θ due to the size of the slit or to that of the blackened area can be made indefinitely small by placing the film far enough behind the slit). This is perfectly consistent with the uncertainty principle because the uncertainty Δp_x represents our *inability to predict* the outcome of an experimental measurement (we cannot tell in advance where on the film the particle will arrive) and *not* the accuracy of the measurement actually made. The second point to note is that the uncertainty principle is about the predictions of the results of two separate sets of experiments. Sometimes one of these is implied rather than performed, because the expected results are so obvious: in the present example, the measurement of x comes into this category. A mistake that is sometimes made is to say that passing the particle through the slit constitutes a 'measurement of position with error Δx', following which its momentum 'is measured with error Δp_x'. However, unless the position of the particle has been recorded, it has not been measured.

Finally, we return to the example of the Heisenberg microscope referred to in chapter 1 (p. 12). As a result of many photons having been scattered by the electron into the microscope we obtain an image of the electron that is 'blurred' by an amount Δx. From the earlier discussion in this section, we see that what this means is that, if we were to follow up this measurement by another which determined x precisely and repeat the whole process a number of times, the spread in the values of x would be Δx. We also derived the uncertainty in the x component of the electron momentum, Δp_x, and we found that these two uncertainties are related by the uncertainty principle as expected. This is correct, but to actually measure Δp_x we would have to conduct a separate series of experiments in which we recorded the particles just after they emerged from the lens. Once again two sets of experiments are required to measure the spread in the results for the two variables.

4.6 The time dependence of the wavefunction

Nearly all our discussion so far has related to the properties of wavefunctions immediately after measurement—i.e., to the eigenfunctions of operators representing physical quantities—and we have given very little consideration to the evolution of the wavefunction in time. However, at the beginning of chapter 2

we set up the time-dependent Schrödinger equation which we generalized to three dimensions in chapter 3 as

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V(\mathbf{r},t)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{4.54}$$

In the notation of the present chapter this becomes

$$\hat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{4.55}$$

where $\hat{H}$ is the Hamiltonian operator for the system. In the earlier chapters we separated off the time dependence and obtained the time-independent Schrödinger equation, which we have since recognized as the energy eigenvalue equation. We shall now postulate that the time-dependent Schrödinger equation (4.55) must always be satisfied even when the energy of the system is not known.

Postulate 4.5 Between measurements, the development of the wavefunction with time is governed by the time-dependent Schrödinger equation.

Note the phrase 'between measurements'. We saw earlier (postulate 4.2) that measurement generally leads to 'collapse' of the wavefunction into one of the eigenfunctions of the measurement operator. This collapse is not a consequence of the Schrödinger equation, but is a separate type of time dependence associated with the act of measurement. The fact that quantum mechanics predicts two different time dependencies for the wavefunction underlies what is known as the quantum measurement 'problem', which will be addressed in the last chapter.

We shall return to a more detailed discussion of the Schrödinger time dependence in chapter 8, but in the meantime we discuss only the particular case where $\hat{H}$ is not itself a function of time; classically this corresponds to a closed system in which energy is conserved. If the energy eigenfunctions of the system obtained by solving the energy eigenvalue equation (4.2) are u_n then, by completeness, the wavefunction at any time t can be expressed as a linear combination of the u_n:

$$\psi(\mathbf{r},t) = \sum_n a_n(t)u_n(\mathbf{r}) \tag{4.56}$$

where the coefficients a_n are, in general, functions of time. Substituting (4.56) into (4.55) we get

$$\begin{aligned} i\hbar\sum_n \frac{da_n}{dt}u_n &= \sum_n a_n\hat{H}u_n \\ &= \sum_n a_n E_n u_n \end{aligned}$$

Thus

$$\sum_n \left(i\hbar\frac{da_n}{dt} - a_n E_n\right)u_n = 0 \tag{4.57}$$

Equation (4.57) must be true at all points in space so the terms in brackets must vanish leading to

$$i\hbar\frac{da_n}{dt} = a_n E_n$$

that is,

$$a_n(t) = a_n(0)\exp(-iE_n t/\hbar) \tag{4.58}$$

where $a_n(0)$ is the value of a_n at some initial time $t = 0$. Hence from (4.56) and (4.58) a general solution to the time-dependent Schrödinger equation is

$$\Psi(\mathbf{r}, t) = \sum_n a_n(0)u_n(\mathbf{r})\exp(-iE_n t/\hbar) \tag{4.59}$$

Now suppose we carry out a measurement of the energy of the system at time $t = 0$. From the earlier postulates we should obtain a result equal to one of the energy eigenvalues, say E_m. Moreover, the wavefunction immediately after the measurement will be the corresponding eigenfunction u_m. This is equivalent to saying that at $t = 0$, $a_m(0) = 1$ and $a_n(0) = 0$, $n \neq m$. In this case (4.59) becomes

$$\Psi(\mathbf{r}, t) = u_m \exp(-iE_m t/\hbar) \tag{4.60}$$

which was the particular form of solution used in chapters 2 and 3. We notice that the right-hand side of (4.60) differs from u_m only by a phase factor, so Ψ is an eigenfunction of $\hat{H}$ at all times (remember $\hat{H}$ is assumed to be time independent). Thus any later measurement of the energy will again yield the value E_m and we conclude that, in this sense, energy is conserved in such a quantum system, as it would be classically. Moreover, the value of any quantity that can be measured compatibly with the energy and whose operator therefore commutes with $\hat{H}$ (e.g., the linear momentum of a free particle) will also be conserved. This result emphasizes the great importance of the time-independent Schrödinger equation (the energy eigenvalue equation) whose solutions we discussed in some detail in earlier chapters. Much of quantum mechanics concerns the properties of the energy eigenstates of systems whose Hamiltonians are independent of time—often referred to as 'stationary states' because of the properties just described. However, many experimental measurements (e.g. atomic spectra) refer to systems which change from one nearly stationary state to another under the influence of a time-dependent potential. We shall discuss the quantum mechanics of such changes in chapter 8 where we shall see that, in many cases, the time-dependent potential can be considered as causing small 'perturbations' on the stationary states of the system.

4.7 Degeneracy

The previous sections of this chapter apply to cases where the eigenfunctions of the operators representing physical measurements are non-degenerate. That is,

we have assumed that if

$$\hat{Q}\phi_n = q_n\phi_n$$

then

$$q_n \neq q_m \qquad \text{for all } n \neq m \tag{4.61}$$

However, degeneracy is in fact quite a common feature of physical systems, and indeed we saw in chapter 3 that it was a necessary consequence of symmetry in a number of three-dimensional examples. In the present section, therefore, we shall extend our discussion of the formal properties of quantum mechanics to the degenerate case. There have been two occasions in which we have explicitly assumed that systems were non-degenerate: first when discussing orthornormality and second when discussing compatible measurements. We shall now extend our discussion of both these points to include degeneracy.

One property that is unique to the degenerate case is that any linear combination of degenerate eigenfunctions (ϕ_n) with the same eigenvalue (q) is also an eigenfunction with that eigenvalue. This follows directly from substitution into (4.61)

$$\begin{aligned}\hat{Q}\sum_n c_n\phi_n &= \sum_n c_n\hat{Q}\phi_n \\ &= q\sum_n c_n\phi_n\end{aligned} \tag{4.62}$$

Turning now to orthogonality, the proof that eigenfunctions *must* be orthogonal no longer holds, but we can show that it is always possible to construct a set of orthogonal eigenfunctions from a set of non-orthogonal eigenfunctions, ϕ_n.

Consider the function ϕ_2' where

$$\phi_2' = S_{12}\phi_1 - \phi_2 \qquad \text{and} \qquad S_{12} = \int \phi_1^*\phi_2\, d\tau$$

then

$$\begin{aligned}\int \phi_1^*\phi_2'\, d\tau &= S_{12}\int \phi_1^*\phi_1\, d\tau - \int \phi_1^*\phi_2\, d\tau \\ &= S_{12} - S_{12} = 0\end{aligned}$$

Thus ϕ_2' is orthogonal to ϕ_1 and a third function orthogonal to these two can be similarly shown to be ϕ_3' where

$$\phi_3' = S_{13}\phi_1 + S_{23}\phi_2' - \phi_3$$

S_{13} is similar to S_{12}, and

$$S_{23} = \int \phi_2'^*\phi_3\, d\tau \Big/ \int |\phi_2'|^2\, d\tau$$

This procedure (known as *Schmidt orthogonalization*) can be continued until a complete orthogonal set of degenerate eigenfunctions has been set up. Thus all the results in this chapter which depend on orthogonality can be extended to the degenerate case, provided it is assumed that such an orthogonal set has first been constructed. Throughout the rest of this book, we shall assume that this has already been done and that all sets of eigenfunctions we encounter are orthogonal.

Turning now to a discussion of compatible measurements, the proof of the theorem that the operators representing compatible measurements commute is clearly unaffected by the presence of degeneracy, but the converse (that measurements represented by commuting operators are compatible) requires modification. Consider the case where ϕ_1 and ϕ_2 are two degenerate eigenfunctions (eigenvalue q) of the operator $\hat{Q}$ which commutes with the operator $\hat{R}$. Assume for the moment that there are only two linearly independent eigenfunctions with this eigenvalue. Then

$$\hat{Q}\hat{R}\phi_1 = \hat{R}\hat{Q}\phi_1 = q\hat{R}\phi_1 \tag{4.63}$$

so $(\hat{R}\phi_1)$ is also an eigenfunction of $\hat{Q}$ with eigenvalue q, and it must therefore be a linear combination of ϕ_1 and ϕ_2. That is,

$$\hat{R}\phi_1 = a\phi_1 + b\phi_2 \tag{4.64}$$

where a and b are constants. An identical argument, replacing ϕ_1 with ϕ_2 in (4.63), leads to

$$\hat{R}\phi_2 = c\phi_1 + d\phi_2 \tag{4.65}$$

where c and d are constants. We shall now show that particular linear combinations of ϕ_1 and ϕ_2 that are eigenfunctions of $\hat{R}$ exist. Let one of these be $\phi' = A\phi_1 + B\phi_2$ and let its eigenvalue be r, then

$$\begin{aligned}\hat{R}\phi' &\equiv A\hat{R}\phi_1 + B\hat{R}\phi_2 \equiv Aa\phi_1 + Ab\phi_2 + Bc\phi_1 + Bd\phi_2 \\ &= r\phi' \equiv rA\phi_1 + rB\phi_2\end{aligned} \tag{4.66}$$

Equating coefficients of ϕ_1 and ϕ_2 in (4.66) we get

$$\left.\begin{aligned}(a-r)A + cB &= 0 \\ bA + (d-r)B &= 0\end{aligned}\right\} \tag{4.67}$$

The equations (4.67) have solutions only if the determinant of the coefficients of A and B vanishes, leading to two possible values of r. In general these will not be equal, so the eigenfunctions of $\hat{R}$ need not be degenerate. This result can readily be extended to the case of an arbitrary number of degenerate functions. We therefore conclude that, although all possible forms of the degenerate eigenfunctions of one operator are not necessarily eigenfunctions of the other, a set of eigenfunctions that are common to both operators can always

be chosen. Once this has been done, compatibility has the same meaning as it did in the non-degenerate case.

Finally we consider the implications of the previous discussion for the quantum theory of measurement. Following a measurement represented by the operator $\hat{Q}$, which produces the result q, the system will have a wavefunction equivalent to one of the set of degenerate eigenfunctions that correspond to this eigenvalue. However, this need not be an eigenfunction of the operator $\hat{R}$ and the exact result of a measurement represented by this operator is not predictable even though $\hat{Q}$ and $\hat{R}$ commute. Nevertheless, the only possible results are those calculable in the manner described earlier, and a subsequent measurement of the quantity represented by $\hat{R}$ will leave q unchanged and result in the wavefunction being equivalent to one of the set of common eigenfunctions. The results of further measurements represented by $\hat{Q}$ or $\hat{R}$ will then be completely predictable.

As an example of this, consider the energy eigenfunctions of the hydrogen atom discussed in chapter 3. A measurement of the energy that showed the atom to have principal quantum number $n = 2$ would result in the atom being in one of four degenerate states ($l = 0$, $m = 0$; $l = 1$, $m = 0$ or ± 1) or in some linear combination of them. A subsequent measurement of the total angular momentum must yield a result $\sqrt{l(l+1)}\hbar$ where l is either zero or one, but which of these is unpredictable. Following this, if l has been found to be equal to one, say, the value of m will be similarly unpredictable unless the z component of angular momentum is also measured. Each of these measurements leaves the values of the previously measured quantities unchanged and, once all three measurements have been made, the wavefunction of the system is completely specified and the results of further measurements of any of these quantities are completely predictable.

4.8 The harmonic oscillator again

We discussed the harmonic oscillator energy states in chapter 2 where we showed that these were quantized with values $(n+\frac{1}{2})\hbar\omega_c$, ω_c being the classical frequency of the oscillator and n an integer $\geqslant 0$. We return to this problem as an example of the application of operator methods, which will lead to the same results. The techniques used here are very similar to those that will be used to analyse the properties of the angular momentum operators (chapter 5) and similar techniques underlie much of quantum field theory, which is touched on in chapter 13.

The energy of a harmonic oscillator is represented by the Hamiltonian operator

$$\hat{H} = \frac{\hat{P}^2}{2m} + \tfrac{1}{2}m\omega_c^2\hat{X}^2 \tag{4.68}$$

We can simplify the algebra in a manner similar to that used in chapter 2 by using dimensionless variables such that x is measured in units of $(\hbar/m\omega_c)^{1/2}$, p in units

of $(m\hbar\omega_c)^{1/2}$ and energy in units of $\hbar\omega_c$ so that (4.68) becomes

$$\hat{H} = \tfrac{1}{2}(\hat{P}^2 + \hat{X}^2) \tag{4.69}$$

We define two new operators in terms of $\hat{X}$ and $\hat{P}$:

$$\begin{aligned} \hat{R}_+ &= (\hat{P} + i\hat{X}) \\ \hat{R}_- &= (\hat{P} - i\hat{X}) \end{aligned} \tag{4.70}$$

In our units the commutation relation (4.41) becomes $[\hat{P}, \hat{X}] = -i$. Using this with (4.70), we get

$$\hat{R}_+\hat{R}_- = (\hat{P}^2 + \hat{X}^2 - i[\hat{P}, \hat{X}]) = 2\hat{H} - 1 \tag{4.71}$$

and

$$\hat{R}_- R_+ = 2\hat{H} + 1 \tag{4.72}$$

so that

$$\hat{H} = \tfrac{1}{2}(\hat{R}_+\hat{R}_- + 1) = \tfrac{1}{2}(\hat{R}_-\hat{R}_+ - 1) \tag{4.73}$$

Using (4.73), we get

$$\begin{aligned} [\hat{R}_+, \hat{H}] &= \hat{R}_+\tfrac{1}{2}(\hat{R}_-\hat{R}_+ - 1) - \tfrac{1}{2}(\hat{R}_+\hat{R}_- + 1)\hat{R}_+ \\ &= -R_+ \end{aligned} \tag{4.74}$$

Similarly

$$[\hat{R}_-, \hat{H}] = \hat{R}_- \tag{4.75}$$

Now, from (4.74)

$$\begin{aligned} \hat{H}\hat{R}_+u_n &= \hat{R}_+\hat{H}u_n + \hat{R}_+u_n \\ &= (E_n + 1)(\hat{R}_+u_n) \end{aligned} \tag{4.76}$$

where u_n is an energy eigenfunction with eigenvalue E_n. Similarly, using (4.75) we get

$$\hat{H}\hat{R}_-u_n = (E_n - 1)(\hat{R}_-u_n) \tag{4.77}$$

Thus $(\hat{R}_+u_n)$ and $(\hat{R}_-u_n)$ are eigenfunctions of $\hat{H}$ with eigenvalues $(E_n + 1)$ and $(E_n - 1)$ respectively. It follows that the operators $\hat{R}_+$ and $\hat{R}_-$ move us up and down a 'ladder' of eigenvalues. As a result $\hat{R}_+$ and $\hat{R}_-$ are known as 'raising' and 'lowering' operators, or sometimes as 'creation' and 'annihilation' operators because they create or annihilate quanta of energy. We also know from the form of the Hamiltonian that all energy levels must be positive. So, if u_0 is the eigenfunction corresponding to the minimum value of the energy, (4.77) can be satisfied only if $\hat{R}_-u_0 = 0$; hence, using (4.71)

$$0 = \hat{R}_+\hat{R}_-u_0 = (2\hat{H} - 1)u_0 = (2E_0 - 1)u_0 \tag{4.78}$$

Hence $E_0 = \frac{1}{2}$ and it follows that the energy levels are $E_n = (n + \frac{1}{2})$ in dimensionless units or

$$E_n = (n + \tfrac{1}{2})\hbar\omega_c \tag{4.79}$$

agreeing with the result in chapter 2.

This is an example of a 'representation-free' derivation, in which the energy eigenvalues have been obtained without using an explicit form of the position and momentum operators.

The form of the eigenfunctions can also be shown to be the same as that deduced in chapter 2. After making the substitutions $\hat{X} = x$, and $\hat{P} = -i\partial/\partial x$, the expression $\hat{R}_- u_0 = 0$ becomes

$$\frac{\partial u_0}{\partial x} + x u_0 = 0 \tag{4.80}$$

whose solution is

$$u_0 = A \exp(-x^2/2) \tag{4.81}$$

where A is a normalizing constant, which again agrees with chapter 2. The wavefunctions of the other states can be obtained by successively operating on u_0 by $\hat{R}_+$, again expressed in terms of x and $-i\partial/\partial x$.

4.9 The measurement of momentum by Compton scattering

We close this chapter with an example of a real physical measurement that illustrates many of the principles discussed here. This is the measurement of the momentum of electrons by Compton scattering of x-rays. We discussed Compton scattering in chapter 1 where we obtained evidence for the existence of photons of energy $\hbar\omega$ and momentum $\hbar\mathbf{k}$, where ω and $\mathbf{k}$ are the angular frequency and wavevector of the x-rays. However, at that stage we assumed that the electron was at rest before the scattering event; we will now lift this restriction and show how Compton scattering can be used to obtain information about the momentum of the scattering electron.

We consider first the case of a free electron whose momentum before and after the photon is scattered has the values $\mathbf{p}$ and $\mathbf{p}'$ respectively and treat this situation, using the classical laws governing the collisions of particles, as we did in chapter 1. Measurement of the wavelength of the incident and scattered x-rays (e.g. by Bragg diffraction from a crystal of known lattice spacing) and knowledge of the direction of the incoming and outgoing photon lead to values for the change in momentum ($\delta\mathbf{p}$) and the change in energy (δE) of the photon in the collision. Using conservation of energy and momentum we have

$$\delta\mathbf{p} = \mathbf{p} - \mathbf{p}' \tag{4.82}$$

and

$$\delta E = p^2/2m - p'^2/2m \tag{4.83}$$

Squaring (4.82) we get

$$p'^2 = p^2 + \delta p^2 - 2\mathbf{p} \cdot \boldsymbol{\delta}\mathbf{p} \tag{4.84}$$

while from (4.83) we have

$$p'^2 = p^2 - 2m\delta E \tag{4.85}$$

Equating the right-hand sides of (4.84) and (4.85) leads to

$$2\mathbf{p} \cdot \boldsymbol{\delta}\mathbf{p} = \delta p^2 + 2m\delta E \tag{4.86}$$

Thus our knowledge of $\delta\mathbf{p}$ and δE have led directly to a measurement of the component of $\mathbf{p}$ in the direction of $\delta\mathbf{p}$ and a similar component of $\mathbf{p}'$ is readily derived from (4.86) and (4.82). We now analyse this measurement from the point of view of quantum mechanics. We first note again that for a free particle ($V = 0$) the momentum operator ($-i\hbar\nabla$) commutes with the Hamiltonian ($-\hbar^2\nabla^2/2m$) so these are compatible measurements and the momentum eigenfunctions are therefore stationary states of the system. The experiment then implies that the electron was in a state of momentum $\mathbf{p}$ (wavefunction $\exp(i\mathbf{p} \cdot \mathbf{r}/\hbar)$) before the measurement and in another stationary state of momentum $\mathbf{p}'$ (wavefunction $\exp(i\mathbf{p}' \cdot \mathbf{r}/\hbar)$) afterwards. The outcome of the experiment is therefore completely predictable and if it is repeated with the same starting conditions, the same result will certainly be obtained for the component of $\mathbf{p}'$ in the direction $\delta\mathbf{p}$.

The situation becomes more complicated, but at the same time more illustrative of the quantum theory of measurement, if the electron is bound to an atom before the scattering process takes place. We shall assume that a previous measurement of the energy of the atom (achieved perhaps by observing photons emitted from it) has shown it to be in its ground state which is, of course, a stationary state of the system. We now consider the dynamics of the scattering process. The conservation of energy and momentum is complicated by the presence of the atomic nucleus and, unless the atom is ionized in the collision, the momentum transfer will be to the atom as a whole and the experiment will give little or no information about the momentum of the individual electron. Even if the collision ionizes the atom, equation (4.86) will provide an accurate measurement of the component of the electron momentum only if the ionization energy is small enough to be ignored compared with the energy transfer, δE: that is if the electron is knocked 'cleanly' out of the atom. Thus in order to measure accurately the momentum of an electron in an atom, we must make the electron effectively free in the process. Looking at this experiment from the point of view of quantum mechanics we first note that, if the electron is bound to the atom, the momentum operator does not commute with the Hamiltonian operator so the initial state cannot be a momentum eigenstate. As we know the change in photon energy, we can use conservation of energy to see that the atom must be in an energy eigenstate after the collision, but if the momentum has been measured

it follows from postulate 4.2 that the electron is also in a momentum eigenstate after the measurement. The only way both these conditions can be fulfilled is if the electron is effectively free after the collision, which is just the condition we derived earlier as necessary for an accurate momentum measurement.

We have seen, therefore, that an accurate measurement of the momentum of an electron in an atom is possible provided the electron is placed in a momentum eigenstate as a result of the experiment. What is not possible on the basis of the postulates of quantum mechanics is a prediction of the precise result of such an experiment. Postulate 4.4, however, tells us how to obtain the relative probabilities of the different possible outcomes and we complete the discussion by applying this to the present example. We assume the atom is hydrogen and that it is initially in its ground state. Then its wavefunction is, from chapter 3,

$$u(\mathbf{r}) = (\pi a_0^3)^{-1/2} \exp(-r/a_0) \tag{4.87}$$

The probability of obtaining a result for $\mathbf{p}$ in the region dp_x, dp_y, dp_z in the vicinity of $\mathbf{p} = \hbar\mathbf{k}$ is $|a(\mathbf{k})|^2$, where $a(\mathbf{k})$ is given by the three-dimensional equivalent of (4.37):

$$\begin{aligned} a(\mathbf{k}) &= \left(\frac{1}{8\pi^3}\right)^{1/2} \int u(\mathbf{r}) \exp(-i\mathbf{k}\cdot\mathbf{r})\, d\tau \\ &= (8\pi^4 a_0^3)^{-1/2} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \exp(-r/a_0 - ikr\cos\theta) r^2\, dr \sin\theta\, d\theta\, d\phi \end{aligned} \tag{4.88}$$

where we have expressed the volume integral in spherical polar coordinates with the direction of $\mathbf{k}$ as the polar axis. The integrals over θ and ϕ are readily evaluated after making the substitution $x = \cos\theta$ leading to

$$f(\mathbf{k}) = 2(2\pi^2 a_0^3)^{-1/2} \int_0^{\infty} k^{-1} \sin(kr) \exp(-r/a_0) r\, dr$$

The integral over r can be evaluated by successive integrations by parts and we obtain

$$f(\mathbf{k}) = 2(2a_0^3)^{1/2}/[\pi(1 + a_0^2 k^2)^2] \tag{4.89}$$

As expected, therefore, the momentum probability distribution for an electron in the ground state of the hydrogen atom is spherically symmetric in momentum space, corresponding with the spherically symmetric position probability distribution obtained by squaring (4.87). Finally we note that the Compton scattering experiment measures one component of momentum only. Calling this the z component, the probability that it will be found to have a value between $\hbar k_z$ and $\hbar(k_z + dk_z)$ while the other components may adopt any values will be

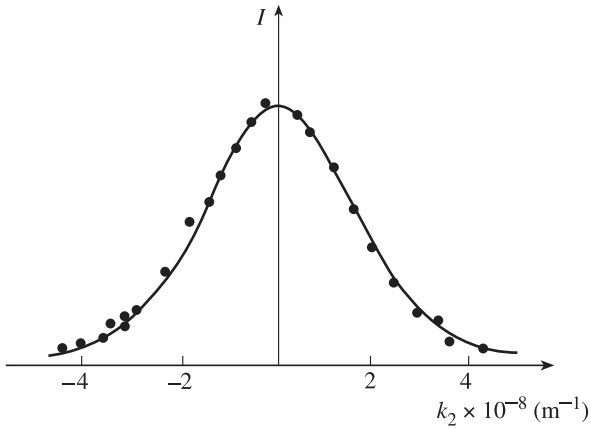


Figure 4.3. A comparison of the theoretical and experimental intensities of Compton scattering from helium. The continuous line represents the momentum probability distribution calculated in the manner described in the text and the points represent measured intensities—after J. W. M. Dumond and H. A. Kirkpatrick, *Physical Review* vol. 52, pp. 419–436, 1937.

$P(k_z)dk_z$ where

$$P(k_z) = \int\int |f(\mathbf{k})|^2 \, dk_x \, dk_y$$
$$= (8a_0^3/\pi^2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1 + a_0^2k_x^2 + a_0^2k_y^2 + a_0^2k_z^2)^{-4} \, dk_x \, dk_y$$

The relevant integrals can be looked up in tables giving

$$P(k_z) = 2a_0/[\pi(1 + a_0^2k_z^2)^2] \tag{4.90}$$

We note that the product of the widths of the momentum distribution ($\delta p_z \simeq \hbar/a_0$) and the position distribution ($\delta z \simeq a_0$) is approximately equal to $\hbar$ in agreement with the uncertainty principle.

We have discussed this example in some detail because it illustrates several of the main features of the basic postulates of quantum mechanics. In particular we have seen how an exact measurement of a physical quantity (in this case the momentum of an electron) results in the system being in an eigenstate of the operator representing the measurement immediately after the measurement has been completed. We have also seen that knowledge of the wavefunction of the system before the measurement does not generally provide a precise prediction of the result, but only of the relative probabilities of the various outcomes.

Another reason for discussing this particular example is that Compton scattering represents a measurement which can be actually carried out.

Unfortunately, such experiments on monatomic hydrogen are impractical, but figure 4.3 shows the momentum component probability distribution in helium, calculated in a similar manner, and also the results of an early Compton scattering experiment, each experimental point representing the number of times the component of momentum of an electron was measured to be in the region of that value. We see that there is excellent agreement between the frequency distribution and the calculated probability distribution, but we note once more that quantum mechanics can only make this statistical prediction and that the values obtained in individual measurements of electron momentum are unpredictable.

Problems

4.1 Show that the operators $\hat{Q}$, $\hat{Q}^2$, $\hat{Q}^3$, etc., are all compatible. Hence show that any component of the momentum of a particle can always be measured compatibly with the kinetic energy, but that the momentum and total energy can be measured compatibly only if the potential energy is constant everywhere.

4.2 A particle moves in one dimension subject to a potential that is zero in the region $-a \leqslant x \leqslant a$ and infinite elsewhere. At a certain time its wavefunction is

$$\psi = (5a)^{-1/2}\cos(\pi x/2a) + 2(5a)^{-1/2}\sin(\pi x/a)$$

What are the possible results of the measurement of the energy of this system and what are their relative probabilities? What are the possible forms of the wavefunction immediately after such a measurement? If the energy is immediately remeasured, what will now be the relative probabilities of the possible outcomes? (The energy eigenvalues and eigenfunctions of this system are given in section 2.4.)

4.3 The energy of the particle in problem 4.2 is measured and a result equal to the lowest energy eigenvalue is obtained. Show that the probability of a subsequent measurement of the electron momentum yielding a result between $\hbar k$ and $\hbar(k + dk)$ is equal to $P(k)\,dk$ where

$$P(k) = \frac{\pi}{2a^3}\frac{\cos^2 ka}{(\pi^2/4a^2 - k^2)^2}$$

4.4 Show that if the particle in problem 4.2 had been in a highly excited energy eigenstate of eigenvalue E, a measurement of the momentum would almost certainly have produced a value equal to $\pm(2mE)^{1/2}$. Compare this result with the predictions of classical mechanics for this problem.

4.5 A particle is observed to be in the lowest energy state of an infinite-sided well. The width of the well is (somehow) expanded to double its size so quickly that the wavefunction does not change *during this process*; the expansion takes place symmetrically so that the centre of the well does not move. The energy of the particle is measured again. Calculate the probabilities that it will be found in (i) the ground state, (ii) the first excited state and (iii) the second excited state of the expanded well.

4.6 Calculate the expectation value of (i) x, (ii) x^2, (iii) p, and (iv) p^2 for a particle known to be in the lowest energy state of a one-dimensional harmonic oscillator potential before these measurements are performed. Hence show that the expectation value of the total energy of the particle equals the ground-state eigenvalue and that the product of the root-mean-square deviations of x and p from their respective means has the minimum value allowed by the uncertainty principle.

4.7 Calculate the expectation values of the following quantities for an electron known to be in the ground state of the hydrogen atom before the measurements are performed: (i) the distance r of the electron from the nucleus, (ii) r^2, (iii) the potential energy, (iv) the kinetic energy. Show that the sum of (iii) and (iv) equals the total energy.

4.8 An operator $\hat{P}$, known as the parity operator, is defined so that, in one dimension, $\hat{P}f(x) = f(-x)$ where f is any well-behaved function of x. Assuming that $\hat{P}$ is real, prove that $\hat{P}$ is Hermitian.

Show that the eigenvalues of $\hat{P}$ are ± 1 and that any function $\phi(x)$ that has definite parity—i.e. if $\phi(x) = \pm\phi(-x)$—is an eigenfunction of $\hat{P}$.

4.9 Show that, if H is the Hamiltonian operator representing the total energy of a system where the potential energy is centrosymmetric and if $\hat{P}$ is the parity operator defined in problem 4.8, then $[\hat{P}, \hat{H}] = 0$. Hence show that in the non-degenerate case, the energy eigenfunctions of such a system always have definite parity. What restrictions, if any, are placed on the eigenfunctions in the similar, but degenerate, case?

4.10 Generalize the results given in problems 4.8 and 4.9 to the three-dimensional case and confirm that they apply to the energy eigenfunctions of a particle in a spherically symmetric potential obtained in chapter 3.

4.11 X-rays of wavelength 1.00×10^{-10} m are incident on a target containing free electrons, and a Compton scattered x-ray photon of wavelength 1.02×10^{-10} m is detected at an angle of 90° to the incident direction. Obtain as much information as possible about the momentum of the scattering electron before and after the scattering process.

Chapter 5

Angular momentum I

Angular momentum is a very important and revealing property of many physical systems. In classical mechanics the principle of conservation of angular momentum is a powerful aid to the solution of such problems as the orbits of planets and satellites and the behaviour of gyroscopes and tops. The role of angular momentum in quantum mechanics is probably even more important and this will be the subject of the present chapter and the next. We shall find that the operators representing the components of angular momentum do not commute with each other, although they all commute with the operator representing the total angular momentum. It follows that no pair of these components can be measured compatibly and we shall therefore look for a set of eigenfunctions that are common to the operators representing the total angular momentum and one of its components. We shall find that the angular momentum eigenvalues always form a discrete set and that, in the case of a central field, the operators commute with the Hamiltonian, implying that the total angular momentum and one component can be measured compatibly with the energy in this case, so that their values, once measured, remain constant in time. Consideration of angular momentum will also enable us to make predictions about the behaviour of atoms in magnetic fields, and we shall find that these are in agreement with experiment only if we assume that the electron (along with other fundamental particles) has an intrinsic 'spin' angular momentum in addition to the 'orbital' angular momentum associated with its motion; the interaction between spin and orbital angular moments leads to detailed features of the atomic spectra which are known as 'fine structure'. We shall also use the measurement of angular momentum to illustrate the quantum theory of measurement discussed in the previous chapter.

Angular momentum is therefore an important subject which deserves and requires detailed consideration. In the rest of this chapter we shall consider the properties of the angular-momentum operators and obtain expressions for their eigenvalues in the case of spin as well as orbital angular momentum, while in the next chapter we shall show that the angular-momentum components can be usefully represented by matrices rather than differential operators, and discuss

how this matrix representation can be used to describe the fine structure of atomic spectra and to illustrate the quantum theory of measurement.

5.1 The angular-momentum operators

The classical expression for the angular momentum $\mathbf{l}$ of a particle whose position and momentum are $\mathbf{r}$ and $\mathbf{p}$ respectively is

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \tag{5.1}$$

Using postulate 4.3, we replace $\mathbf{r}$ and $\mathbf{p}$ by the operators $\hat{\mathbf{R}}$ and $\hat{\mathbf{P}}$. The quantum-mechanical operator, $\hat{\mathbf{L}}$, representing angular momentum is then

$$\hat{\mathbf{L}} = \hat{\mathbf{R}} \times \hat{\mathbf{P}} \tag{5.2}$$

where $\hat{\mathbf{R}}$ and $\hat{\mathbf{P}}$ are the operators representing position and momentum. The operators representing the Cartesian components of angular momentum are therefore

$$\hat{L}_x = \hat{Y}\hat{P}_z - \hat{Z}\hat{P}_y \qquad \hat{L}_y = \hat{Z}\hat{P}_x - \hat{X}\hat{P}_z \qquad \hat{L}_z = \hat{X}\hat{P}_y - \hat{Y}\hat{P}_x \tag{5.3}$$

and that representing the square of the magnitude of the total angular momentum is

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \tag{5.4}$$

We now investigate whether these operators represent quantities that are compatible in the sense discussed in the previous chapter. To do this, we check whether the operators $\hat{L}_x$, $\hat{L}_y$, $\hat{L}_z$, and $\hat{L}^2$ commute when taken in pairs. Using (5.3), we get

$$\begin{aligned}
[\hat{L}_x, \hat{L}_y] &= \hat{L}_x\hat{L}_y - \hat{L}_y\hat{L}_x \\
&= (\hat{Y}\hat{P}_z - \hat{Z}\hat{P}_y)(\hat{Z}\hat{P}_x - \hat{X}\hat{P}_z) - (\hat{Z}\hat{P}_x - \hat{X}\hat{P}_z)(\hat{Y}\hat{P}_z - \hat{Z}\hat{P}_y) \\
&= \hat{Y}\hat{P}_z\hat{Z}\hat{P}_x - \hat{Y}\hat{P}_z\hat{X}\hat{P}_z - \hat{Z}\hat{P}_y\hat{Z}\hat{P}_x + \hat{Z}\hat{P}_y\hat{X}\hat{P}_z \\
&\quad - \hat{Z}\hat{P}_x\hat{Y}\hat{P}_z + \hat{Z}\hat{P}_x\hat{Z}\hat{P}_y + \hat{X}\hat{P}_z\hat{Y}\hat{P}_z - \hat{X}\hat{P}_z\hat{Z}\hat{P}_y
\end{aligned} \tag{5.5}$$

The corresponding terms in the last two lines of (5.5) contain the same operators, but in a different order, which is important only for those operators that do not commute. Remembering the commutation relations for the components of position and momentum (4.42), the only non-commuting operators in (5.5) are $\hat{P}_z$ and $\hat{Z}$. We can therefore cancel the second and third terms in the second last line with the corresponding terms in the last line, and rearrange the others to give

$$\begin{aligned}
[\hat{L}_x, \hat{L}_y] &= \hat{Y}\hat{P}_x(\hat{P}_z\hat{Z} - \hat{Z}\hat{P}_z) + \hat{X}\hat{P}_y(\hat{Z}\hat{P}_z - \hat{P}_z\hat{Z}) \\
&= (\hat{X}\hat{P}_y - \hat{Y}\hat{P}_x)(\hat{Z}\hat{P}_z - \hat{P}_z\hat{Z}) \\
&= i\hbar(\hat{X}\hat{P}_y - \hat{Y}\hat{P}_x) \\
&= i\hbar\hat{L}_z
\end{aligned} \tag{5.6}$$

using (5.3). $[\hat{L}_y, \hat{L}_z]$ and $[\hat{L}_z, \hat{L}_x]$ can be obtained similarly and we have

$$[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z \tag{5.7}$$

$$[\hat{L}_y, \hat{L}_z] = i\hbar\hat{L}_x \tag{5.8}$$

$$[\hat{L}_z, \hat{L}_x] = i\hbar\hat{L}_y \tag{5.9}$$

We can now complete our calculation of the commutator brackets by considering those involving $\hat{L}^2$. From (5.4)

$$[\hat{L}^2, \hat{L}_z] = [\hat{L}_x^2, \hat{L}_z] + [\hat{L}_y^2, \hat{L}_z] + [\hat{L}_z^2, \hat{L}_z] \tag{5.10}$$

Considering the first term on the right-hand side and using equations (5.7)–(5.9)

$$\begin{aligned}[\hat{L}_x^2, \hat{L}_z] &= \hat{L}_x\hat{L}_x\hat{L}_z - \hat{L}_z\hat{L}_x\hat{L}_x \\ &= \hat{L}_x(-i\hbar\hat{L}_y + \hat{L}_z\hat{L}_x) - (i\hbar\hat{L}_y + \hat{L}_x\hat{L}_z)\hat{L}_x \\ &= -i\hbar(\hat{L}_x\hat{L}_y + \hat{L}_y\hat{L}_x)\end{aligned} \tag{5.11}$$

Similarly

$$[\hat{L}_y^2, \hat{L}_z] = i\hbar(\hat{L}_x\hat{L}_y + \hat{L}_y\hat{L}_x) \tag{5.12}$$

Also

$$[\hat{L}_z^2, \hat{L}_z] = \hat{L}_z^3 - \hat{L}_z^3 = 0 \tag{5.13}$$

Substituting from (5.11), (5.12), and (5.13) into (5.10) and generalizing the result to the commutator brackets containing the other components we get

$$[\hat{L}^2, \hat{L}_x] = [\hat{L}^2, \hat{L}_y] = [\hat{L}^2, \hat{L}_z] = 0 \tag{5.14}$$

It follows from (5.7)–(5.9) that the operators representing any two components of angular momentum do not commute and are therefore not compatible. If, therefore, the system is in an eigenstate of one angular momentum component, it will not be simultaneously in an eigenstate of either of the others. However, (5.14) shows that the total angular momentum *can* be measured compatibly with any one component. We shall therefore approach the eigenvalue problem by looking for a set of functions which are simultaneously eigenfunctions of the total angular momentum and *one* component, which is conventionally chosen to be the z component $\hat{L}_z$.

5.2 The eigenvalues and eigenfunctions

We can obtain an explicit expression for the operators representing the angular momentum components using the forms of the position and momentum operators given in postulate 4.3:

$$\hat{\mathbf{P}} = -i\hbar\nabla \qquad \hat{\mathbf{R}} = \mathbf{r}$$

Hence

$$\hat{\mathbf{L}} = -i\hbar\mathbf{r} \times \nabla \tag{5.15}$$

When considering the angular momentum of a particle moving about a point, it is a great advantage to refer to a spherical polar coordinate system similar to that used in chapter 3 when discussing the energy eigenvalues of a particle in a central field. Expressions for $\hat{L}_z$ and $\hat{L}^2$ in this coordinate system will be derived using vector calculus, but readers who are not familiar with this technique may prefer to accept equations (5.18) and (5.19) and proceed to the discussion following these equations.

We first refer the vector $\mathbf{r}$ to the origin of the spherical polar coordinate system (cf. figure 3.2), giving us the standard expression

$$\nabla = \mathbf{r}_0\frac{\partial}{\partial r} + \frac{1}{r}\boldsymbol{\theta}\frac{\partial}{\partial\theta} + \frac{1}{r\sin\theta}\boldsymbol{\phi}\frac{\partial}{\partial\phi} \tag{5.16}$$

where $\mathbf{r}_0$, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are the three basic unit vectors of the spherical polar system. It follows directly that

$$\hat{\mathbf{L}} = -i\hbar\mathbf{r} \times \nabla = -i\hbar\left(\boldsymbol{\phi}\frac{\partial}{\partial\theta} - \frac{1}{\sin\theta}\boldsymbol{\theta}\frac{\partial}{\partial\phi}\right) \tag{5.17}$$

Taking the polar axis to be z, the unit vector, $\mathbf{z}_0$, in this direction is

$$\mathbf{z}_0 = \cos\theta\mathbf{r}_0 - \sin\theta\boldsymbol{\theta}$$

Hence

$$\hat{L}_z = \mathbf{z}_0 \cdot \hat{\mathbf{L}} = -i\hbar\frac{\partial}{\partial\phi} \tag{5.18}$$

An expression for $\hat{L}^2$ is obtained from the vector identity

$$\begin{aligned}\hat{L}^2 &= -\hbar^2(\mathbf{r} \times \boldsymbol{\nabla}) \cdot (\mathbf{r} \times \boldsymbol{\nabla}) \\ &= -\hbar^2\mathbf{r} \cdot [\boldsymbol{\nabla} \times (\mathbf{r} \times \boldsymbol{\nabla})]\end{aligned}$$

Substituting for $\mathbf{r} \times \boldsymbol{\nabla}$ from (5.17) this expression becomes

$$\hat{L}^2 = -\hbar^2\mathbf{r} \cdot \boldsymbol{\nabla} \times \left(\boldsymbol{\phi}\frac{\partial}{\partial\theta} - \boldsymbol{\theta}\frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right)$$

so that, using the standard expression for curl in spherical polar coordinates we get

$$\hat{L}^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right] \tag{5.19}$$

We can now use (5.18) and (5.19) to obtain the eigenvalues and the common set of eigenfunctions for the operators $\hat{L}_z$ and $\hat{L}^2$. Considering $\hat{L}^2$

first, we notice that (5.19) is identical (apart from a multiplicative constant) to the differential equation (3.26) determining the angularly dependent part of the energy eigenfunctions in the central field problem which has already been solved in chapter 3. It follows directly from (3.40) that the eigenvalues of $\hat{L}^2$ are

$$l(l+1)\hbar^2 \tag{5.20}$$

where l is a positive integer or zero, and that the corresponding eigenfunctions—cf. (3.44)—are

$$Y_{lm}(\theta,\phi) = (-1)^m \left[\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} P_l^{|m|}(\cos\theta)e^{im\phi} \tag{5.21}$$

where m is an integer whose modulus is less than or equal to l, and some of the properties of the spherical harmonics Y_{lm} and the associated Legendre functions, $P_l^{|m|}$, have been discussed in chapter 3. It follows immediately from (5.18) and (5.21) that

$$\hat{L}_z Y_{lm} = -i\hbar\frac{\partial Y_{lm}}{\partial\phi} = m\hbar Y_{lm} \tag{5.22}$$

so the functions Y_{lm} are the simultaneous eigenfunctions of $\hat{L}_z$ and $\hat{L}^2$ that we have been looking for and the eigenvalues of $\hat{L}_z$ are given by

$$m\hbar \qquad \text{where} \qquad -l \leqslant m \leqslant l \tag{5.23}$$

We have therefore confirmed the physical interpretation of the quantum numbers l and m that we anticipated in chapter 3: the square of the total angular momentum is quantized in units of $l(l+1)\hbar^2$ and the z component in units of $m\hbar$. The condition $-l \leqslant m \leqslant l$ corresponds to the obvious requirement that a measurement of the total angular momentum must yield a result at least as large as that from a simultaneous measurement of any component. If the magnitude of the total angular momentum and that of the z component are known, it follows that the angular momentum vector must be at a fixed angle to the z axis. However, its orientation with respect to the x and y axes is unknown and neither of these components can therefore be measured compatibly with $\hat{L}_z$, which is consistent with the results derived earlier from the commutation relations. The possible orientations with respect to the z axis are illustrated in figure 5.1 for the case where $l = 2$: we note that there is no eigenstate where the angular momentum is exactly parallel to the z axis, which is to be expected as this configuration would imply that the x and y components were simultaneously known to be exactly equal to zero.

Our knowledge of the angular-momentum eigenvalues and eigenfunctions also helps us to understand more clearly the physical basis of the degeneracy found in the energy states of the central-field problem in chapter 3. The central-field Hamiltonian, which is contained in the time-independent Schrödinger

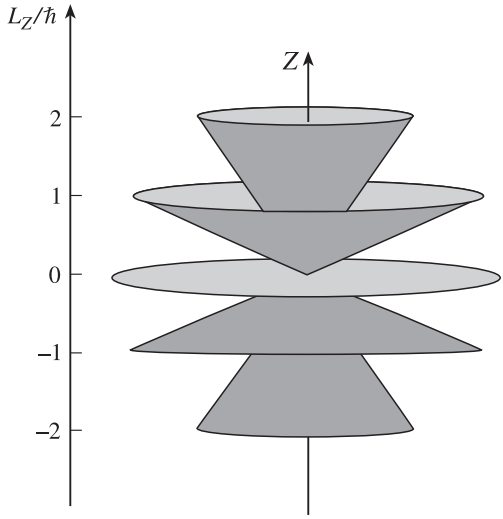


Figure 5.1. The possible orientations of the angular momentum vector relative to the z axis in the case where $l = 2$. Note that the orientation within the xy plane is unknown if the z component has a definite value, so the angular momentum vector joins the origin to the edge of a cone whose axis is z.

equation (3.23), can be written using (5.19) in the form

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + V(r) + \frac{\hat{L}^2}{2mr^2} \tag{5.24}$$

The final term on the right-hand side is the quantum-mechanical equivalent of the centrifugal term in the classical expression for the total energy. The only part of $\hat{H}$ which depends on θ and ϕ is this last term so it follows from this, and the fact that $\hat{L}^2$ and $\hat{L}_z$ are not functions of r, that $\hat{H}$ commutes with both $\hat{L}^2$ and $\hat{L}_z$, so that the total energy eigenfunctions are, in this case, also eigenfunctions of the total angular momentum and of one component. Energy levels corresponding to the same value of l, but different values of m are degenerate, so the angular part of the wavefunction may be any linear combination of the eigenfunctions corresponding to these various values of m. This is equivalent to saying that the spatial orientation of the angular momentum vector is completely unknown so that no component of it can be measured unless the spherical symmetry of the problem is broken—for example by applying a magnetic field as will be discussed in the next section.

5.3 The experimental measurement of angular momentum

Now that we have derived expressions for the angular momentum eigenvalues, we shall see how these agree with the results of experimental measurement. Such measurements are often made by studying the effects of magnetic fields on the motion of particles and we shall consider these in this section.

We first treat the problem classically. We consider the classical motion of a particle, such as an electron, with charge $-e$ and mass m_e moving in a circle with angular velocity ω in an orbit of radius r. Its angular momentum $\mathbf{l}$ therefore has magnitude $m_e\omega r^2$ and points in a direction perpendicular to the plane of the orbit. The circulating charge is clearly equivalent to an electric current $-e\omega/2\pi$ moving round a loop of area πr^2. Its magnetic moment $\boldsymbol{\mu}$ has magnitude equal to the product of the current and the loop area which is $e\omega r^2/2$. Its direction is parallel, but opposite to $\mathbf{l}$ (because of the negative sign). Hence we have

$$\boldsymbol{\mu} = -\frac{e}{2m_e}\mathbf{l} \tag{5.25}$$

We can now apply postulate 4.3 to (5.25) to obtain the quantum-mechanical operator representing a measurement of magnetic moment:

$$\hat{\boldsymbol{\mu}} = -\frac{e}{2m_e}\hat{\mathbf{L}}$$

that is,

$$\hat{\mu}_z = -\frac{e}{2m_e}\hat{L}_z \qquad \text{etc.} \tag{5.26}$$

It follows that, if the z component of the magnetic moment of an atom is measured, a value for this component of the angular momentum is also obtained.

If a magnetic field $\mathbf{B}$ is applied to such a system, the energy of interaction between the magnetic moment and the field will be $-\boldsymbol{\mu}\cdot\mathbf{B}$. The corresponding quantum-mechanical operator in the case where $\mathbf{B}$ is in the z direction is therefore $\Delta\hat{H}$ where

$$\Delta\hat{H} = \frac{eB}{2m_e}\hat{L}_z \tag{5.27}$$

and, if the system is in an eigenstate of $\hat{L}_z$ with eigenvalue $m\hbar$, a measurement of this interaction energy will yield the value

$$\Delta E = \mu_B B m \tag{5.28}$$

where

$$\mu_B = \frac{e\hbar}{2m_e} \tag{5.29}$$

and is known as the *Bohr magneton*. Thus, for example, if we apply a magnetic field to an atom in a p state (i.e. one with $l = 1, m = -1, 0$ or 1) the interaction energy (5.28) will lift its threefold degeneracy. We would therefore

expect spectral lines resulting from a transition between this state and a non-degenerate s ($l = 0, m = 0$) state to be split into a triplet, the angular frequency difference between neighbouring lines being equal to $(eB/2m_e)$. However, when this experiment is performed on a one-electron atom such as hydrogen, a rather different result is observed. For strong fields the spectra are consistent with the $2p$ level having been split into *four* substates instead of three and the $1s$ level, which was expected to remain single, into *two*. As the latter state possesses no angular momentum associated with the motion of the electron in the field of the nucleus, the observed magnetic moment and associated angular momentum must arise from some other cause.

As we shall show, these results can be explained if we postulate that an electron possesses an additional *intrinsic* angular momentum in addition to the *orbital* angular momentum discussed so far. Following normal practice, we shall often refer to the intrinsic angular momentum as 'spin', but, as will be discussed later, this should not be taken to mean that the electron is actually spinning. In the $l = 0$ state this intrinsic angular momentum can apparently adopt two orientations, while in the $l = 1$ state the orbital and intrinsic angular momenta couple to produce the four substates observed. The details of this 'spin–orbit coupling' will be discussed in the next chapter.

The Stern–Gerlach experiment

We have seen that the energy of an atom which possesses a magnetic moment is changed by the interaction between this magnetic moment and an applied magnetic field. If this applied field is inhomogeneous—i.e. varies from place to place—there will be a force on the atom directing it towards regions where the interaction energy is smaller. Thus, if an atom with a z component of magnetic moment μ_z is in a magnetic field directed along z whose magnitude B is a function of z, it will be subject to a force F given by

$$F = -\mu_z \frac{\partial B}{\partial z} \tag{5.30}$$

An experiment which makes use of this effect to measure atomic magnetic moments is the *Stern–Gerlach* experiment. A beam of atoms is directed between two poles of a magnet that are shaped as shown in figure 5.2(b) in order to produce an inhomogeneous field which exerts a force on the atoms in a direction transverse to the beam. On emerging from the field, the atoms will have been deflected by an amount proportional to this component of their magnetic moment. In the absence of spin, therefore, we should expect the beam to be split into $2l + 1$ parts where l is the total orbital angular momentum quantum number and, in the particular case where l is zero, we should then expect no splitting at all. However, when this experiment is performed on hydrogen (or a similar one-electron system) in its ground state, the atomic beam is found to be split into two, implying that the atom possesses a magnetic moment whose z component can adopt two opposite

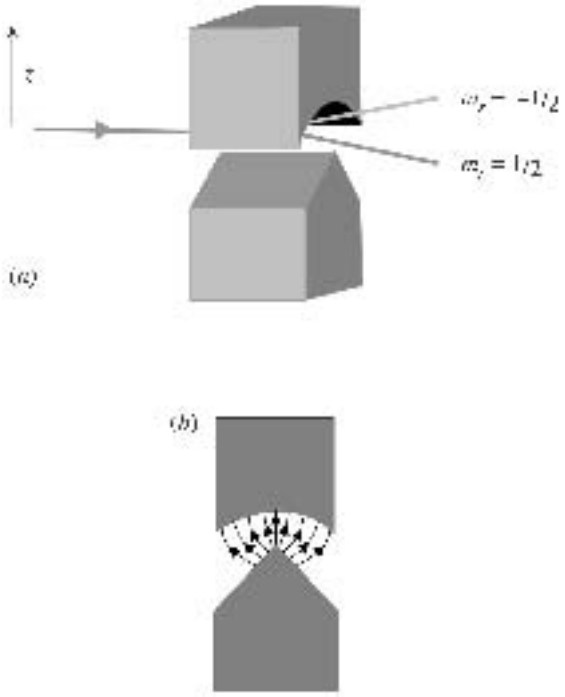


Figure 5.2. (*a*) A beam of spin-half atoms is split by an inhomogeneous magnetic field. The form of the field is shown in projection down the beam direction in (*b*).

orientations with respect to the field. As the ground state of hydrogen has zero angular momentum, this magnetic moment cannot be associated with the orbital motion of the electron and must be due to its spin. Thus the Stern–Gerlach experiment provides direct evidence for the electron possessing intrinsic angular momentum, which is consistent with the spectroscopic results described earlier. It is clear, however, that the rules governing the quantization of spin must be different from those governing orbital angular momentum, as the latter require there to be an odd number $(2l + 1)$ of eigenstates of L_z whereas an even number (2) is observed in the Stern–Gerlach experiment. This would be consistent only if the total angular momentum quantum number l equalled one half in this case, with the quantum number representing the z component equal to $\pm\frac{1}{2}$. But our earlier discussion, based on the analysis in chapter 3, showed that these should be integers.

5.4 General solution to the eigenvalue problem

If we re-examine the arguments leading to the quantization of orbital angular momentum, which were set out when the equivalent problem of the energy eigenfunctions of a particle in a central field was discussed in chapter 3, we see that it resulted from the boundary condition on the wavefunction requiring it to be a continuous, single-valued, integrable function of the particle position. However, spin is not associated with the motion of the particle in space and its eigenfunctions are therefore not functions of particle position, so the same conditions need not apply in this case. However, spin is a form of angular momentum so it might be reasonable to expect the quantum-mechanical operators representing its components to obey the same algebra as do the corresponding operators in the orbital case. In particular, we might expect the commutation relations (5.7)–(5.9) and (5.14) to be universally valid for all forms of angular momentum. We shall assume this to be the case and, using only this assumption and the basic physical condition that the total angular momentum eigenvalues must be at least as large as those of any simultaneously measured component, we shall obtain expressions for the angular-momentum eigenvalues that are valid for spin, as well as in the orbital case.

Let the simultaneous eigenvalues of $\hat{L}^2$ and $\hat{L}_z$ be α and β respectively and let the corresponding eigenfunction be ϕ, where ϕ is not necessarily a function of particle position. (A suitable representation in the case of spin will be described in the next chapter.) We then have

$$\hat{L}^2\phi = \alpha\phi \qquad \text{and} \qquad \hat{L}_z\phi = \beta\phi \qquad \text{where } \alpha \geqslant \beta^2 \tag{5.31}$$

We now define two operators $\hat{L}_+$ and $\hat{L}_-$ as

$$\hat{L}_+ = \hat{L}_x + i\hat{L}_y \qquad \hat{L}_- = \hat{L}_x - i\hat{L}_y \tag{5.32}$$

These operators do not represent physical quantities (note that they are not Hermitian) but are very useful in developing the mathematical argument that follows. For reasons that will soon become clear they are known as *raising and lowering operators*, or sometimes as *ladder operators*. Similar operators were used in the treatment of the harmonic oscillator in chapter 4.

We first establish some commutation relations involving $\hat{L}_+$ and $\hat{L}_-$. From (5.32), using (5.4) and (5.7),

$$\begin{aligned}\hat{L}_+\hat{L}_- &= \hat{L}_x^2 + \hat{L}_y^2 - i[\hat{L}_x, \hat{L}_y] \\ &= \hat{L}^2 - \hat{L}_z^2 + \hbar\hat{L}_z\end{aligned} \tag{5.33}$$

Similarly

$$\hat{L}_-\hat{L}_+ = \hat{L}^2 - \hat{L}_z^2 - \hbar\hat{L}_z \tag{5.34}$$

Hence

$$[\hat{L}_+, \hat{L}_-] = 2\hbar\hat{L}_z \tag{5.35}$$

We also have

$$\begin{aligned}[\hat{L}_z, \hat{L}_+] &= [\hat{L}_z, \hat{L}_x] + i[\hat{L}_z, \hat{L}_y] \\ &= i\hbar(\hat{L}_y - i\hat{L}_x)\end{aligned}$$

using (5.8) and (5.9). That is,

$$[\hat{L}_z, \hat{L}_+] = \hbar\hat{L}_+ \tag{5.36}$$

and similarly,

$$[\hat{L}_z, \hat{L}_-] = -\hbar\hat{L}_- \tag{5.37}$$

Having established these relations, which we shall refer to again shortly, we now operate on both sides of the second of equations (5.31) with $\hat{L}_+$ giving

$$\hat{L}_+\hat{L}_z\phi = \hat{L}_+\beta\phi$$

Using (5.36) and remembering that β is a constant, this leads to

$$\hat{L}_z(\hat{L}_+\phi) = (\beta + \hbar)(\hat{L}_+\phi) \tag{5.38}$$

Similarly, using (5.31) and (5.37), we get

$$\hat{L}_z(\hat{L}_-\phi) = (\beta - \hbar)(\hat{L}_-\phi) \tag{5.39}$$

Thus, if ϕ is an eigenfunction of $\hat{L}_z$ with eigenvalue β, $(\hat{L}_+\phi)$ and $(\hat{L}_-\phi)$ are also eigenfunctions of $\hat{L}_z$ with eigenvalues $(\beta + \hbar)$ and $(\beta - \hbar)$ respectively.

We also have, from the first of the equations (5.31),

$$\hat{L}_+\hat{L}^2\phi = \alpha\hat{L}_+\phi \qquad \text{and} \qquad \hat{L}_-\hat{L}^2\phi = \alpha\hat{L}_-\phi \tag{5.40}$$

Now we know that $\hat{L}^2$ commutes with both $\hat{L}_x$ and $\hat{L}_y$ so it must also commute with $\hat{L}_+$ and $\hat{L}_-$; (5.40) therefore becomes

$$\hat{L}^2(\hat{L}_+\phi) = \alpha(\hat{L}_+\phi) \qquad \text{and} \qquad \hat{L}^2(\hat{L}_-\phi) = \alpha(\hat{L}_-\phi) \tag{5.41}$$

so the functions $(\hat{L}_+\phi)$ and $(\hat{L}_-\phi)$ as well as being eigenfunctions of $\hat{L}_z$ are also eigenfunctions of $\hat{L}^2$ with eigenvalue α. It follows that the operators $\hat{L}_+$ and $\hat{L}_-$ respectively 'raise' and 'lower' the eigenfunctions up or down the 'ladder' of eigenvalues of $\hat{L}_z$ corresponding to the same eigenvalue of $\hat{L}^2$, so accounting for the names of the operators mentioned earlier.[1]

We can now apply the condition that β^2 is less than or equal to α which implies that there must be both a maximum (say β_1) and minimum (say β_2) value

[1] The operators $\hat{L}_+$ and $\hat{L}_-$ are also known as 'creation' and 'annihilation' operators respectively because they 'create' or 'annihilate' quanta of $\hat{L}_z$.

of β corresponding to a particular value of α. If ϕ_1 and ϕ_2 are the corresponding eigenfunctions, we can satisfy (5.38) and (5.39) only if

$$\hat{L}_+\phi_1 = 0 \tag{5.42}$$

and

$$\hat{L}_-\phi_2 = 0 \tag{5.43}$$

We now operate on (5.42) with $\hat{L}_-$ and use (5.34) to get

$$\hat{L}_-\hat{L}_+\phi_1 = (\hat{L}^2 - \hat{L}_z^2 - \hbar\hat{L}_z)\phi_1 = 0 \tag{5.44}$$

and hence, using (5.31),

$$(\alpha - \beta_1^2 - \hbar\beta_1)\phi_1 = 0$$

That is,

$$\alpha = \beta_1(\beta_1 + \hbar) \tag{5.45}$$

Similarly, from (5.43), (5.33) and (5.31),

$$\alpha = \beta_2(\beta_2 - \hbar) \tag{5.46}$$

It follows from (5.45) and (5.46) (remembering that β_2 must be less than β_1 by definition) that

$$\beta_2 = -\beta_1 \tag{5.47}$$

Now we saw from (5.38) and (5.39) that neighbouring values of β are separated by $\hbar$ so it follows that β_1 and β_2 are separated by an integral number (say n) of steps in $\hbar$. That is,

$$\beta_1 - \beta_2 = n\hbar \tag{5.48}$$

It follows directly from (5.47) and (5.48) that

$$\beta_1 = -\beta_2 = n\hbar/2$$

and hence, putting n equal to $2l$ and using (5.45)

$$\alpha = l(l+1)\hbar^2 \qquad \text{where } l = n/2 \tag{5.49}$$

and

$$\beta = m\hbar \tag{5.50}$$

where m varies in integer steps between $-l$ and l, and l is either an integer *or a half-integer*.

We see that this result is exactly the one we have been looking for. In the case of orbital angular momentum, the extra condition that the wavefunction must be a well-behaved function of particle position requires n to be even and l to be an integer, so that (5.49) and (5.50) are equivalent to (5.20) and (5.23). However,

when we are dealing with intrinsic angular momentum this condition no longer holds and if n is odd, l and m are half-integers. In the particular case of electron spin, the Stern–Gerlach experiment showed that the z-component operator has two eigenstates, implying that the total-spin quantum number equals one-half in this case, with $m = \pm\frac{1}{2}$. This is always true for the electron and also for other fundamental particles such as the proton, the neutron and the neutrino which are accordingly known as 'spin-half' particles. Other particles exist with total-spin quantum numbers that are integers or half-integers greater than one-half, and their angular-momentum properties are also exactly as predicted by this theory.

The possession by the electron of two kinds of angular momentum suggests an analogy with the classical case of a planet, such as the earth, orbiting the sun: it has orbital angular momentum associated with this motion and intrinsic angular momentum associated with the fact that it is spinning about an axis. Because of this, the intrinsic angular momentum of an electron (or other particle) is usually referred to as *spin*. However, there is no evidence to suggest that the electron is literally spinning and, indeed, it turns out that such a mechanical model could account for the observed angular momentum only if the electron had a structure, parts of which would have to be moving at a speed faster than that of light, so contradicting the theory of relativity! In fact the only satisfactory theory of the origin of spin does result from a consideration of relativistic effects: Dirac showed that quantum mechanics and relativity could be made consistent only if the electron is assumed to be a point particle which possesses *intrinsic* angular momentum. By 'intrinsic' we mean that this is a fundamental property of the electron, just like its charge or mass. Although conventionally referred to as 'spin', nothing is assumed about its underlying structure. We shall give a brief introduction to Dirac's theory in chapter 11, where we shall see that all fundamental particles, such as the electron, should have intrinsic angular momentum with quantum number equal to one-half. Experimentally, it is found that particles that are not spin-half[2] are either 'exchange' particles such as the photon which are not subject to Dirac's theory, or they have a structure, being composed of several more fundamental spin-half particles. For example, the alpha particle has total spin zero, but this results from a cancellation of the spins of the constituent protons and neutrons. The truly fundamental constituents of matter (apart from the photon and other 'exchange' particles that are not subject to Dirac's theory) are believed to be *leptons* (which include the electron) and *quarks* and these are all spin-half particles.

In order to proceed further with our study of angular momentum, we shall have to consider the eigenfunctions as well as the eigenvalues of the angular-

[2] It is very important when referring to particles as 'spin-half', 'spin-one' etc. to remember that these labels refer to the total-spin quantum number and not to the magnitude of the angular momentum which actually has the value $\sqrt{3}\hbar/2$ in the first case and $\sqrt{2}\hbar$ in the second. Similarly an electron with z component of spin equal to $\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$ is often referred to as having 'spin up' or 'spin down' respectively even though its angular momentum vector is inclined at an angle of $\cos^{-1}(3^{-1/2}) \simeq 55°$ to the z axis—cf. figure 5.1.

momentum operators. However, we have already seen that if this discussion is to include spin, these eigenfunctions cannot be expressed as functions of particle position. In the next chapter, therefore, we shall develop an alternative representation of quantum mechanics based on matrices, which will be used to describe the properties of intrinsic angular momentum.

Normalization

We have seen that the operators $\hat{L}_+$ and $\hat{L}_-$ operate on the eigenfunction ϕ_m of $\hat{L}_z$ to produce the eigenfunctions ϕ_{m+1} and ϕ_{m-1}. However, there is nothing in the earlier treatment to ensure that the eigenfunctions so created are normalized and, in general, they are not. To ensure normalization, we have to multiply the eigenfunctions by appropriate constants, after creating them using the ladder operators, as we now show.

We first show that, although $\hat{L}_+$ and $\hat{L}_-$ are not Hermitian operators, they are *Hermitian conjugates* in the sense that

$$\int f\hat{L}_+ g\, d\tau = \int g\hat{L}_-^* f\, d\tau \tag{5.51}$$

This follows straightforwardly from the definitions of $\hat{L}_+$ and $\hat{L}_-$ (5.32) and the fact that $\hat{L}_x$ and $\hat{L}_y$ are Hermitian:

$$\int f\hat{L}_+ g\, d\tau = \int f(\hat{L}_x + i\hat{L}_y)g\, d\tau = \int g(\hat{L}_x + i\hat{L}_y)^* f\, d\tau = \int g\hat{L}_-^* f\, d\tau$$

We now show how a normalized eigenfunction, ϕ_{m+1}, can be generated from ϕ_m, given that the latter is normalized.

$$\begin{aligned}\int \phi_{m+1}\phi_{m+1}^* d\tau &= \int (\hat{L}_+\phi_m)\hat{L}_+^*\phi_m^* d\tau \\ &= \int \phi_m^*\hat{L}_-\hat{L}_+\phi_m d\tau = [l(l+1) - m(m+1)]\hbar\end{aligned} \tag{5.52}$$

using (5.51), (5.44) and the fact that ϕ_m is normalized. To produce a ladder of normalized eigenfunctions, we must therefore use

$$\phi_{m+1} = [l(l+1) - m(m+1)]^{-1/2}\hbar^{-1}\hat{L}_+\phi_m \tag{5.53}$$

Similarly in the case of $\hat{L}_-$,

$$\phi_{m-1} = [l(l+1) - m(m-1)]^{-1/2}\hbar^{-1}\hat{L}_-\phi_m \tag{5.54}$$

These results will be used during the general treatment of the coupling angular momenta toward the end of the next chapter.

Problems

5.1 A point mass μ is attached by a massless rigid rod of length a to a fixed point in space and is free to rotate in any direction. Find a classical expression for the energy of the system if its total angular momentum has magnitude L, and hence show that the quantum-mechanical energy levels are given by $l(l+1)\hbar^2/2ma^2$ where l is a positive integer.

5.2 The mass and rod in problem 5.1 are now mounted so that they can rotate only about a particular axis that is perpendicular to the rod and fixed in space. What are the energy levels now? What can be said about the components of angular momentum perpendicular to the axis of such a system when it is in an energy eigenstate?

5.3 A particle has a wavefunction of the form $z\exp[-\alpha(x^2+y^2+z^2)]$ where α is a constant. Show that this function is an eigenfunction of $\hat{L}^2$ and $\hat{L}_z$ and find the corresponding eigenvalues. Use the raising and lowering operators L_+ and L_- to obtain (unnormalized) expressions for all the other eigenfunctions of L_z that correspond to the same eigenvalue of $\hat{L}^2$.

Hint: Use Cartesian coordinates.

5.4 Verify directly that the spherical harmonics Y_{lm} with $l \leqslant 2$ (see chapter 3) are eigenfunctions of $\hat{L}^2$ and $\hat{L}_z$ with the appropriate eigenvalues.

5.5 The total orbital angular momentum of the electrons in a silver atom turns out to be zero while its total-spin quantum number is $\frac{1}{2}$, resulting in a magnetic moment whose z component has a magnitude of 1 Bohr magneton (9.3×10^{-24} J T^{-1}). In the original Stern–Gerlach experiment, silver atoms, each with a kinetic energy of about 3×10^{-20} J travelled a distance of 0.03 m through a non-uniform magnetic field of gradient 2.3×10^3 T m^{-1}. Calculate the separation of the two beams 0.25 m beyond the magnet.

5.6 The wavefunction of a particle is known to have the form $f(r,\theta)\cos\phi$. What can be predicted about the likely outcome of a measurement of the z component of angular momentum of this system?

5.7 Derive (5.54) in a similar manner to that set out for (5.53).

5.8 Express $\hat{L}_x$ and $\hat{L}_y$ in terms of $\hat{L}_+$ and $\hat{L}_-$ and hence show that for a system in an eigenstate of $\hat{L}_z$, $\langle\hat{L}_x\rangle = \langle\hat{L}_y\rangle = 0$. Also obtain expressions for $\langle\hat{L}_x^2\rangle$ and $\langle\hat{L}_y^2\rangle$ (using (5.53) and (5.54) where necessary) and compare the product of these two quantities with the predictions of the uncertainty principle.

5.9 Obtain expressions for $\hat{L}_x$ and $\hat{L}_y$ in spherical polar coordinates and hence show that

$$\hat{L}_\pm = \hbar e^{\pm i\phi}\left(\pm\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\phi}\right)$$

5.10 Use the raising and lowering operators to derive expressions for all the spherical harmonics with $l=2$, given that

$$Y_{20} = \left(\frac{5}{16\pi}\right)^{1/2}(3\cos^2\theta - 1)$$

Compare your results with those given in chapter 3.

Chapter 6

Angular momentum II

In the first part of this chapter we shall show how it is possible to represent dynamical variables by matrices instead of differential operators without affecting the predicted results of physically significant quantities. We shall see that although this representation is often more complicated and cumbersome than the methods used earlier, it has a particularly simple form when applied to problems involving angular momentum, and has the great advantage that it can be used in the case of spin where no differential operator representation exists. We shall use such spin matrices to analyse experiments designed to measure the component of spin in various directions and find that these provide an important and illustrative example of the quantum theory of measurement.

In the second part of this chapter we shall discuss the problem of the addition of different angular momenta (such as the orbital and spin angular momenta of an electron in an atom) and illustrate the results by discussing the effects of spin–orbit coupling and applied magnetic fields on the spectra of one-electron atoms.

6.1 Matrix representations

We first consider the case of a dynamical variable represented by the Hermitian operator $\hat{Q}$. If one of its eigenvalues is q and the corresponding eigenfunction is ψ then

$$\hat{Q}\psi = q\psi \tag{6.1}$$

We now expand ψ in terms of some complete orthonormal set of functions ϕ_n, which are not necessarily eigenfunctions of $\hat{Q}$—cf. (4.29):

$$\psi = \sum_n a_n \phi_n \tag{6.2}$$

As in chapter 4, we assume that the ϕ_n form a discrete set, but the extension to the continuous case is reasonably straightforward. Substituting from (6.2) into (6.1)

we get

$$\sum_n a_n \hat{Q}\phi_n = q \sum_n a_n \phi_n \tag{6.3}$$

We now multiply both sides of (6.3) by the complex conjugate of one of the functions—say ϕ_m^*—and integrate over all space. This leads to

$$\sum_n Q_{mn} a_n = q a_m \tag{6.4}$$

where the Q_{mn} are defined by

$$Q_{mn} = \int \phi_m^* \hat{Q} \phi_n \, d\tau \tag{6.5}$$

and we have used the orthogonality property in deriving the right-hand side of (6.4). Equation (6.4) is true for all values of m so we can write it in matrix form:

$$\begin{bmatrix} Q_{11} & Q_{12} & \cdots \\ Q_{21} & Q_{22} & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = q \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \tag{6.6}$$

where it is clear from (6.5) and the fact that $\hat{Q}$ is a Hermitian operator, that $Q_{mn}^* = Q_{nm}$—cf. (4.13). A matrix whose elements obey this condition is known as a *Hermitian matrix*, and equation (6.6) is a *matrix eigenvalue equation*. Such an equation has solutions only for the particular values of q that satisfy the determinantal condition

$$\begin{vmatrix} Q_{11} - q & Q_{12} & \cdots \\ Q_{21} & Q_{22} - q & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{vmatrix} = 0 \tag{6.7}$$

There are as many solutions to this equation as there are rows (or columns) in the determinant. As we have already defined q as one of the eigenvalues of the original operator $\hat{Q}$, it follows that the eigenvalues of the matrix equation (6.6) are the same as those of the original differential equation (6.1) and these two equations are therefore formally equivalent. Equation (6.6) can therefore be written in the form

$$[Q][a] = q[a] \tag{6.8}$$

where $[Q]$ is the matrix whose elements are Q_{mn} and $[a]$ is the column vector whose elements are a_n. In the same way that ψ is an eigenfunction of the Hermitian *operator* $\hat{Q}$, the column vector $[a]$ is an 'eigenvector' of the Hermitian matrix $[Q]$.

The properties of Hermitian matrices are closely analogous to those of Hermitian operators discussed in chapter 4. To demonstrate this, we often need to transpose a matrix (by exchanging its rows and columns) or a vector (by replacing a column vector by a row vector and vice versa) and at the same time take its complex conjugate. This process is known as 'taking the Hermitian conjugate' of the matrix or vector and is symbolized by a † superscript. Thus, for any matrix $[A]$

$$[A^\dagger] \equiv [\tilde{A}^*]$$

so that if $[Q]$ is a Hermitian matrix

$$[Q^\dagger] = [Q] \tag{6.9}$$

For example, we can show that the eigenvalues of a Hermitian matrix are real—cf. (4.18). We take the Hermitian conjugate of (6.8)

$$[a^\dagger][Q] = q^*[a^\dagger] \tag{6.10}$$

using the standard rule for transposing a product of matrices and (6.9). We now pre-multiply (6.8) by $[a^\dagger]$ to get

$$[a^\dagger][Q][a] = q[a^\dagger][a]$$

and post-multiply (6.10) by $[a]$ to get

$$[a^\dagger][Q][a] = q^*[a^\dagger][a]$$

from which it follows that $q = q^*$ and the eigenvalue is real.

We now consider orthonormality which was expressed in chapter 4 as

$$\int \phi_n^* \phi_m \, d\tau = \delta_{nm} \tag{6.11}$$

If $[a']$ is another eigenvector of $[Q]$ with eigenvalue q', we post-multiply (6.10) by $[a']$ to get

$$[a^\dagger][Q][a'] = q^*[a^\dagger][a'] \tag{6.12}$$

We now pre-multiply (6.8) by $[a'^\dagger]$ to get

$$[a'^\dagger][Q][a] = q'[a'^\dagger][a]$$

and take the Hermitian conjugate of this equation:

$$[a^\dagger][Q][a'] = q'^*[a^\dagger][a'] \tag{6.13}$$

The left-hand sides of (6.12) and (6.13) are equal, so it follows that

$$[a^\dagger][a'] = 0 \tag{6.14}$$

unless $q = q'$. By orthogonality of eigenvectors we therefore mean that the scalar product of one with the Hermitian conjugate of the other is zero. We can also ensure normalization, in the sense that $[a^\dagger][a] = 1$, by multiplying all the elements of the eigenvector by an appropriate constant. The other properties of Hermitian operators set out in chapter 4 can be converted to matrix form in a similar manner, the main difference being that where integrals of products of operators and functions appear in the former case, they are replaced by products of matrices and vectors in the latter.

We see from this that the physical properties of a quantum-mechanical system can be derived using appropriate matrix equations instead of the differential operator formalism. The postulates and algebra contained in chapter 4 can all be expressed directly in matrix terms: the dynamical variables are represented by Hermitian matrices and the state of the system is described by a 'state vector' that is identical to the appropriate eigenvector immediately after a measurement whose result was equal to the corresponding eigenvalue. Moreover, provided we could postulate appropriate matrices to represent the dynamical variables, the whole procedure could be carried out without ever referring to the original operators or eigenfunctions.

However, the matrix method suffers from one major disadvantage. This is that because most quantum systems have an infinite number of eigenstates, the dynamical variables, including position and momentum, usually have to be represented by matrices of infinite order. Techniques for obtaining the eigenvalues in some such cases have been developed and, indeed, Heisenberg developed a form of quantum mechanics based on matrices (sometimes known as 'matrix mechanics') and used it to solve the energy eigenvalue equations for the simple harmonic oscillator and the hydrogen atom before 'wave mechanics' had been invented by Schrödinger. Nevertheless, matrices of infinite order are generally difficult to handle and solving the corresponding differential equations is nearly always easier. This problem does not arise when the number of eigenvalues is finite, which accounts for the usefulness of matrix methods when studying angular momentum. Provided the total angular momentum has a fixed value (determined by the quantum number l) only the $2l + 1$ states with different values of the quantum number m have to be considered and the angular momentum operators can be represented by matrices of order $2l + 1$. This simplification is particularly useful in the case of spin, as for spin-half particles, $s = \frac{1}{2}$ (s is the equivalent of l in the case of spin) and therefore 2×2 matrices can be used. Moreover, we have seen that it is not possible to represent the spin components by differential operators so matrix methods are particularly useful in this case.

6.2 Pauli spin matrices

A suitable set of 2×2 matrices which can be used to represent the angular-momentum components of a spin-half particle were first discovered by W. Pauli

and are known as *Pauli spin matrices*. These are:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{6.15}$$

and the operators $\hat{S}_x$, $\hat{S}_y$ and $\hat{S}_z$ representing the three spin components are expressed in matrix form as:

$$\hat{S}_x = \tfrac{1}{2}\hbar\sigma_x \qquad \text{etc.} \tag{6.16}$$

To see that these matrices do indeed form a suitable representation for spin, we first check whether they obey the correct commutation relations:

$$\begin{aligned} [\hat{S}_x, \hat{S}_y] &= \tfrac{1}{4}\hbar^2\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} - \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \\ &= \tfrac{1}{4}\hbar^2\left(\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} - \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}\right) \\ &= i\tfrac{1}{2}\hbar^2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= i\hbar S_z \end{aligned} \tag{6.17}$$

This result agrees with that derived earlier (5.7) and the other commutation relations can be similarly verified. Turning now to the eigenvalues, we see that those of $\hat{S}_z$ are simply $\frac{1}{2}\hbar$ times the eigenvalues of σ_z which are in turn obtained from the 2×2 equivalent of (6.7):

$$\begin{vmatrix} 1-q & 0 \\ 0 & -1-q \end{vmatrix} = 0 \tag{6.18}$$

Hence $q = \pm 1$ and the eigenvalues of $\hat{S}_z$ are $\pm\frac{1}{2}\hbar$ as expected. The corresponding eigenvectors are obtained from the eigenvalue equation, which in the case $q = +1$ is

$$\frac{1}{2}\hbar \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{1}{2}\hbar \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{6.19}$$

which is satisfied if $a_2 = 0$; if we require the eigenvector to be normalized, we also have $a_1 = 1$. In the case $q = -1$, the corresponding results are $a_1 = 0$ and $a_2 = 1$. We therefore conclude that the eigenvectors corresponding to the eigenvalues $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ respectively. Moreover, these are clearly orthogonal as expected from (6.14).

Eigenvalues and eigenvectors can similarly be found for $\hat{S}_x$ and $\hat{S}_y$. In each case the eigenvalues are equal to $\pm\frac{1}{2}\hbar$ and the corresponding eigenvectors are shown in table 6.1. The reader should verify that each eigenvector is normalized and that each pair is orthogonal.

Table 6.1. The eigenvalues and eigenvectors of the matrices representing the angular momentum components of a spin-half particle. NB: The overall phase of an eigenvector, like that of a wavefunction, is arbitrary.

Spin component	Eigenvalue	Eigenvector
$\hat{S}_x$	$\frac{1}{2}\hbar$	$\alpha_x \equiv \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$
$\hat{S}_x$	$-\frac{1}{2}\hbar$	$\beta_x \equiv \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$
$\hat{S}_y$	$\frac{1}{2}\hbar$	$\alpha_y \equiv \frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$
$\hat{S}_y$	$-\frac{1}{2}\hbar$	$\beta_y \equiv \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$
$\hat{S}_z$	$\frac{1}{2}\hbar$	$\alpha_z \equiv \begin{bmatrix}1\\0\end{bmatrix}$
$\hat{S}_z$	$-\frac{1}{2}\hbar$	$\beta_z \equiv \begin{bmatrix}0\\1\end{bmatrix}$

As a further test of the correctness of this representation, we consider the square of the total angular momentum whose operator is given by

$$\begin{aligned}\hat{S}^2 &= \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2 \\ &= \frac{3}{4}\hbar^2\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\end{aligned} \tag{6.20}$$

using (6.15) and (6.16). Clearly $\hat{S}^2$ commutes with $\hat{S}_x$, $\hat{S}_y$ and $\hat{S}_z$ as expected and all the eigenvectors given in table 6.1 are also eigenvectors of $\hat{S}^2$ with eigenvalue $\frac{1}{2}(\frac{1}{2}+1)\hbar^2$. We conclude therefore that the Pauli spin matrix representation for the angular-momentum components of a spin-half particle produces the same results as were obtained previously, and we can proceed with confidence to use this representation to obtain further information about the properties of spin.

6.3 Spin and the quantum theory of measurement

The measurement of spin provides a very clear illustration of the quantum theory of measurement. Consider a beam of spin-half particles travelling along the y axis, as in figure 6.1, towards a Stern–Gerlach apparatus oriented to measure the z component of spin (such an apparatus will be denoted by SGZ). We assume that the number of particles in each beam is small enough for there to be only one in

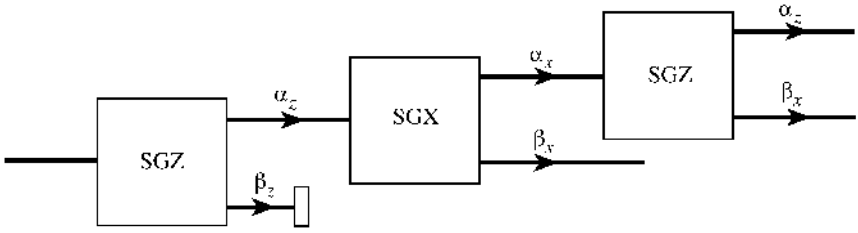


Figure 6.1. Successive measurements of the angular momentum components of spin-half particles. The boxes represent sets of Stern–Gerlach apparatus which direct a particle into the upper or lower output channel depending on whether the appropriate spin component is found to be positive or negative. The y direction is from left to right.

the Stern–Gerlach apparatus at any one time; we can therefore treat each particle independently and ignore interactions between them. Two beams of particles will emerge, one with z component of spin equal to $-\frac{1}{2}\hbar$, which we block off with a suitable stop, and the other with z component $\frac{1}{2}\hbar$, which we allow to continue. This measurement has therefore defined the state of the system and the state vector of the particles that carry on is known to be the eigenvector α_z of $\hat{S}_z$, corresponding to an eigenvalue of $\frac{1}{2}\hbar$ (cf. table 6.1). If we were to pass this beam directly through another SGZ apparatus, we should expect all the particles to emerge in the channel corresponding to $S_z = \frac{1}{2}\hbar$ and none to emerge in the other channel, and indeed this prediction is confirmed experimentally.

We now consider the case where the particles that emerge from the first SGZ with a positive z component pass into a similar apparatus oriented to measure the x component of spin (that is an SGX) as in figure 6.1. Each particle must emerge from the SGX with S_x equal to either $\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$, but, as the initial state vector is not an eigenvector of $\hat{S}_x$, we cannot predict which result will actually be obtained. However, we can use the quantum theory of measurement (postulate 4.4) to predict the relative probabilities of the possible outcomes. We expand the initial state vector [1] as a linear combination of the eigenvectors of $\hat{S}_x$ and the probabilities are then given by the squares of the appropriate expansion coefficients. Thus, referring to table 6.1,

$$\alpha_z = c_+\alpha_x + c_-\beta_x \tag{6.21}$$

[1] To define the complete state of the particle we should multiply the vector representing the spin state of the particle by a wavefunction representing its motion in space. For example, the full state of the particle emerging from the first SGZ apparatus would have the form $\exp(iky)\begin{bmatrix}1\\0\end{bmatrix}$. It simplifies the treatment if this is implied rather than explicitly included.

where c_+ and c_- are constants. That is,

$$\begin{bmatrix}1\\0\end{bmatrix} = c_+\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} + c_-\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix} \tag{6.22}$$

from which it follows directly that $c_+ = c_- = 2^{-1/2}$ and therefore

$$\alpha_z = \frac{1}{\sqrt{2}}(\alpha_x + \beta_x) \tag{6.23}$$

Thus the two possible results of a measurement of the x component of spin have equal probability and, although we cannot predict the result of the measurement on any particular particle, we know that if the experiment is repeated a large number of times, half the particles will emerge in each channel. We saw in chapter 4 that an important aspect of quantum measurement theory is that the act of measurement leaves the system in an eigenstate of the measurement operator, but destroys knowledge of any other property that is not compatible with it. To emphasize this point, consider one of the beams of particles emerging from the SGX to be passed through a second SGZ (see figure 6.1 again); a similar argument to the previous one shows that we shall again be unable to predict in which channel a particular particle will emerge, although the relative probabilities of the two possible results are again equal.

We see therefore that the act of measuring one component of spin has destroyed any knowledge we previously had about the value of another. In fact the quantum theory of measurement (at least as it is conventionally interpreted) states that while a particle is in an eigenstate of $\hat{S}_z$, it is meaningless to think of it as having a value for the x component of its spin, as the latter can be measured only by changing the state of the system to be an eigenstate of $\hat{S}_x$. Moreover, as we have seen, the result of a measurement of the x component is completely unpredictable if the particle is in an eigenstate of $\hat{S}_z$, and this indeterminacy is to be thought of as an intrinsic property of such a quantum system. Some scientists have found these features of quantum mechanics very unsatisfactory and attempts have been made to develop other theories, known as 'hidden variable theories', that avoid the problems of indeterminacy and the like. We shall discuss these in some detail in chapter 13 where we shall show that, where such theories predict results different from those of quantum mechanics, experiment has always favoured the latter. In that chapter we shall also discuss the measurement problem in more depth, again making use of the measurement of spin components as an illustrative example. These discussions will involve the consideration of the measurement of two spin components that are not necessarily at right angles to each other and, as this also provides another example of the application of Pauli spin matrices, we shall derive some relevant results for this case at this point.

We consider a modification of the set-up illustrated in figure 6.1 in which the second Stern–Gerlach apparatus (SGϕ) is oriented to measure the spin component in the xz plane at an angle ϕ to z. The operator $\hat{S}_\phi$, representing the component

of angular momentum in this direction, is given by analogy with the classical expression as

$$\begin{aligned}\hat{S}_\phi &= \hat{S}_z \cos\phi + \hat{S}_x \sin\phi \\ &= \frac{1}{2}\hbar \begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{bmatrix}\end{aligned} \tag{6.24}$$

using (6.15). The eigenvalues of this matrix are easily shown to be $\pm\frac{1}{2}\hbar$ as expected and if the eigenvector in the positive case is $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$, then

$$\tfrac{1}{2}\hbar(a_1 \cos\phi + a_2 \sin\phi) = \tfrac{1}{2}\hbar a_1$$

and therefore

$$\frac{a_1}{a_2} = \frac{\sin\phi}{1-\cos\phi} = \frac{\cos(\phi/2)}{\sin(\phi/2)} \tag{6.25}$$

The normalized eigenvector is therefore $\begin{bmatrix} \cos(\phi/2) \\ \sin(\phi/2) \end{bmatrix}$ and a similar argument leads to the expression $\begin{bmatrix} -\sin(\phi/2) \\ \cos(\phi/2) \end{bmatrix}$ in the case where the eigenvalue equals $-\frac{1}{2}\hbar$. We note that when $\phi = 0$, the eigenvectors are the same as those of $\hat{S}_z$; and when $\phi = \pi/2$ they are identical to those of $\hat{S}_x$ (cf. table 6.1); this is to be expected as the ϕ direction is parallel to z and x respectively in these cases. The probabilities of obtaining a positive or negative result from such a measurement are $|c_+|^2$ and $|c_-|^2$ respectively, where c_+ and c_- are now the coefficients of the SGϕ eigenvectors in the expansion of the initial state vector. Thus

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_+ \begin{bmatrix} \cos(\phi/2) \\ \sin(\phi/2) \end{bmatrix} + c_- \begin{bmatrix} -\sin(\phi/2) \\ \cos(\phi/2) \end{bmatrix} \tag{6.26}$$

from which it follows that

$$c_+ = \cos(\phi/2) \qquad \text{and} \qquad c_- = -\sin(\phi/2) \tag{6.27}$$

and the probabilities of a positive or negative result are therefore $\cos^2(\phi/2)$ and $\sin^2(\phi/2)$ respectively. We note that when $\phi = 0$ the result is certain to be positive because in this case the second apparatus is identical to the first, and that when $\phi = \pi/2$ (corresponding to a measurement of the x component) the probability of each is equal to one-half in agreement with the earlier result. Experiments have been performed to carry out measurements such as those described earlier, and in each case they produce results in agreement with those predicted by the quantum theory of measurement. We shall discuss such experiments in more detail in chapter 13 where we shall find that the confirmation of results such as (6.27) is an important test of the correctness of quantum mechanics when compared with 'hidden variable' theories.

6.4 Dirac notation

P. A. M. Dirac was born in France and spent his academic career in the University of Cambridge from the 1930s until his death in 1984. He is best known for his work in developing the 'Dirac equation' which extends quantum mechanics into the high-energy regime where relativistic effects are very important (see chapter 12). Two main consequences of this are the fact that electrons and other fundamental particles automatically acquire an intrinsic angular momentum or 'spin' and that every such particle also has an 'antiparticle'—e.g. the positron is the antiparticle of the electron.

Another major contribution by Dirac was to put the whole of quantum mechanics on a more general footing. He was one of the first to realize that the measurable results of the theory are independent of the 'basis' used to describe them. By this we mean that alternative formulations, such as the wavefunction and any matrix representation, are completely equivalent, as far as their predictions of experimental results are concerned. Dirac developed a 'representation-free' notation to express this. Consider a typical quantum expression such as

$$Q_{12} = \int \psi_1^* \hat{Q} \psi_2 d\tau \tag{6.28}$$

in wave notation and its equivalent in matrix notation which is

$$Q_{12} = [a_1][Q][a_2] \tag{6.29}$$

In Dirac notation we write

$$Q_{12} = \langle 1|\hat{Q}|2\rangle \tag{6.30}$$

as a general expression that can stand for either of the other two. To express an eigenfunction equation such as

$$\hat{Q}\phi_n = q_n\phi_n \qquad \text{or} \qquad [Q][a]_n = q[a]_n \tag{6.31}$$

in Dirac notation, we split the 'bracket' in (6.30) into a 'bra' $\langle n|$ and a 'ket' $|n\rangle$ to get

$$\hat{Q}|n\rangle = q|n\rangle \tag{6.32}$$

We note that what is written inside the bra or the ket is quite arbitrary and can be chosen to suit the context of the problem. For example, the ground state of the hydrogen atom might be written as

$$|n = 1, l = 0, m = 0, s = \tfrac{1}{2}\rangle$$

or

$$|1, 0, 0, \tfrac{1}{2}\rangle$$

or even as

$$|\text{the ground state of the hydrogen atom}\rangle.$$

Dirac developed this notation when he formulated quantum mechanics in the language of 'vector spaces'. The bras and kets form coordinate systems for two abstract 'Hilbert' spaces of, in general, infinite dimensions. These ideas underlie most advanced thinking about quantum mechanics, but we shall confine ourselves to using Dirac notation as a notation that can be translated into whatever wave or matrix representation is most useful for the problem in hand. To recover the wave representation we should replace the ket by a wavefunction, and the bra by its complex conjugate. If they appear together as in (6.30) the integration (6.28) or matrix product (6.29) is implied. In the rest of this book, we shall use Dirac notation sparingly, but an example of its application will be found in the section on combining angular momenta towards the end of this chapter.

6.5 Spin–orbit coupling and the Zeeman effect

In the last chapter we showed that, classically, an electron with orbital angular momentum $\mathbf{l}$ has a magnetic moment $\boldsymbol{\mu}_l$ given by

$$\boldsymbol{\mu}_l = -\frac{e}{2m_e}\mathbf{l} \tag{6.33}$$

and that the corresponding quantum-mechanical operators $\hat{\boldsymbol{\mu}}_l$ and $\hat{\mathbf{L}}$ are similarly related. There is a similar magnetic moment $\boldsymbol{\mu}_s$ associated with electron spin which is proportional to the spin angular momentum **(s)**, but the constant of proportionality in this case cannot be easily calculated and must be obtained from experiment or from relativistic quantum theory; its value turns out to be almost exactly twice that in the orbital case and we therefore have

$$\boldsymbol{\mu}_s = -\frac{e}{m_e}\mathbf{s} \tag{6.34}$$

The orbital and spin magnetic moments separately interact with any applied magnetic field (the *Zeeman effect*) and also with each other (*spin–orbit coupling*). To obtain an expression for the quantum-mechanical operator representing the spin–orbit energy, we follow the same procedure as before and calculate an expression for the corresponding classical quantity, which we then quantize following the procedure set out in chapter 4. We first consider the particular case of a one-electron atom, such as hydrogen, and extend our treatment to the more general case later.

Classically, an electron orbiting a nucleus of charge Ze 'sees' the nucleus in orbit around it. Assuming the orbit to be circular and the angular velocity ω, the electron is at the centre of a loop carrying a current equal to $Ze\omega/2\pi$ and is therefore subject to a magnetic field of magnitude

$$B = \frac{\mu_0 Ze\omega}{4\pi r} = \frac{\mu_0 Ze}{4\pi m_e r^3} l \tag{6.35}$$

where l is, as before, the magnitude of the classical angular momentum of the electron—i.e. $l = m_e r^2 \omega$. This is an example of a more general expression for the magnetic field $\mathbf{B}$ seen by a particle moving with velocity $\mathbf{v}$ in an electric field $\mathcal{E}(\mathbf{r})$ which is

$$\mathbf{B} = -\mathbf{v} \times \mathcal{E}/c^2 \tag{6.36}$$

where c is the speed of light. If the potential V associated with the field $\mathbf{E}$ is spherically symmetric (as in a one-electron atom) we have

$$\mathcal{E} = -\frac{\partial V}{\partial r}\frac{\mathbf{r}}{r} \tag{6.37}$$

and hence

$$\begin{aligned} \mathbf{B} &= \frac{\mathbf{v} \times \mathbf{r}}{rc^2}\frac{\partial V}{\partial r} \\ &= -\frac{1}{m_e c^2 r}\frac{\partial V}{\partial r}\mathbf{l} \end{aligned} \tag{6.38}$$

In the case of a one-electron atom $V(r) = Ze/(4\pi\epsilon_0 r)$ and (6.38) is the same as (6.35).

If the magnetic moment associated with the electron spin is $\boldsymbol{\mu}_s$ the energy of interaction (W) between it and this magnetic field is given by

$$W = -\tfrac{1}{2}\boldsymbol{\mu}_s \cdot \mathbf{B} \tag{6.39}$$

where the factor of one-half arises from a relativistic effect known as Thomas precession, which we shall not discuss further here except to note that it has nothing to do with the other relativistic factor of one-half mentioned in connection with equation (6.34). Combining (6.34), (6.35) and (6.39), we get

$$W = \frac{\mu_0 Z e^2}{8\pi m_e^2 r^3}\mathbf{l} \cdot \mathbf{s} \tag{6.40}$$

If $\mathbf{j}$ is the total angular momentum $\mathbf{l} + \mathbf{s}$ then

$$j^2 = l^2 + s^2 + 2\mathbf{l} \cdot \mathbf{s} \tag{6.41}$$

and hence, combining (6.40) and (6.41)

$$W = \frac{\mu_0 Z e^2}{16\pi m_e^2 r^3}(j^2 - l^2 - s^2) \tag{6.42}$$

We can make the transformation from classical mechanics to quantum mechanics in the usual way by replacing the dynamical variables with appropriate operators. In the presence of an external magnetic field of magnitude B_0, which

we assume to be in the z direction, the operator representing the interaction energy between the spin moment and both the orbital and applied fields is $\hat{H}'$ where

$$\hat{H}' = f(r)(\hat{J}^2 - \hat{L}^2 - \hat{S}^2) - \frac{eB_0}{2m_e}(\hat{L}_z + 2\hat{S}_z) \tag{6.43}$$

and

$$f(r) = \frac{\mu_0 Z e^2}{4\pi m_e^2 r^3} = \frac{-e}{m_e^2 c^2 r}\frac{\partial V}{\partial r} \tag{6.44}$$

We will shortly discuss the effect of this on the energy of a one-electron atom. We shall need to remember some of the properties of the hydrogen-atom energy eigenfunctions we derived in chapter 3. In particular, those eigenfunctions are also eigenfunctions of the squared total orbital angular momentum $\hat{L}^2$ (eigenvalue $l(l+1)\hbar^2$) and its z component $\hat{L}_z$ (eigenvalue $m_l\hbar$); we also note that eigenfunctions with the same l and different m_l are degenerate.

A general solution to (6.43) is a little complicated to derive, so we shall first treat some special cases. Some readers may prefer to proceed directly to the general treatment in section 6.6.

6.5.1 The strong-field Zeeman effect

The most straightforward case is where the second term in (6.43) is much larger than the first, which is known as *the strong-field Zeeman Effect*. This means that the effect of the applied field is much greater than the spin–orbit coupling, which we ignore to a first approximation. In this case, simple products of eigenfunctions of $\hat{L}_z$ and $\hat{S}_z$ are clearly also eigenfunctions of $\hat{H}'$. The energy eigenvalues are therefore changed by an amount equal to ΔE, where

$$\Delta E = \frac{e\hbar B}{2m_e}(m_l + 2m_s) \tag{6.45}$$

and m_l and m_s are the quantum numbers associated with the z components of the orbital and spin angular momenta. Each state corresponding to given values of l and s is therefore split into $(2l+1)(2s+1)$ equally-spaced components, although some of these may be accidentally degenerate. For example, if $l = 1$, the possible values of m_l are ± 1 or 0, while those of m_s, are $\pm\frac{1}{2}$. Hence, $(m_l + 2m_s)$ can equal ± 2, ± 1 or 0 where the last state is doubly degenerate.

Although for typical high field strengths the spin–orbit interaction is smaller than that representing the interaction with the field, it may not be so small that it can be ignored completely. A further correction, δE, to each state specified by m_l and m_s can be estimated by assuming that the eigenfunctions are the same as in the absence of any spin–orbit interaction and the changes to the eigenvalues are just the expectation values of the spin–orbit contribution to the Hamiltonian (this

is an example of perturbation theory to be discussed in the next chapter). We get

$$\begin{aligned}\delta E &= 2\langle f(r)\rangle\langle \hat{\mathbf{L}}\cdot\hat{\mathbf{S}}\rangle \\ &= 2\langle f(r)\rangle m_l m_s \hbar^2 \end{aligned} \tag{6.46}$$

where we have used the fact that $\langle \hat{L_x}\rangle = \langle \hat{L_y}\rangle = \langle \hat{S_x}\rangle = \langle \hat{S_y}\rangle = 0$ when the system is in an eigenstate of $\hat{L_z}$ and $\hat{S_z}$.

6.5.2 Spin–orbit coupling

We now turn to the opposite extreme where the applied field is zero and we can ignore the second term of (6.43). The first point to note is that if $B_0 = 0$, there is nothing to define the direction of the z axis and therefore no necessity for the energy eigenfunctions to be eigenfunctions of $\hat{L_z}$ or $\hat{S_z}$. However, the first term contains $\hat{J}^2$, the squared magnitude of the total angular momentum. The energy eigenfunctions must therefore be simultaneous eigenfunctions of $\hat{J}^2$, $\hat{L}^2$ and $\hat{S}^2$, which can, in general, be expressed as linear combinations of the eigenfunctions of $\hat{L_z}$ and $\hat{S_z}$. At the end of the last chapter, we showed from very general considerations that the energy levels of any total angular momentum operator have the form $j(j+1)\hbar^2$, where j is an integer or a half integer. It follows from this and (6.43) that the changes in the energy levels are proportional to

$$[j(j+1) - l(l+1) - s(s+1)]\hbar^2$$

The constant of proportionality can again be estimated by assuming that the radial part of the wavefunction is unaffected by $\hat{H}'$ so that the energy change, δE can be approximated by its expectation value. (This will be more rigorously justified in the next chapter.) We therefore have

$$\delta E = \langle f(r)\rangle[j(j+1) - l(l+1) - s(s+1)]\hbar^2 \tag{6.47}$$

The allowed values of j for given l and s can be deduced from the following physical argument, which will be made more rigorous later. We consider the two extreme cases where the orbital angular momenta are (i) as closely parallel and (ii) as closely antiparallel as possible, given the restrictions on j, l, and s. In the first case, $j = l + s$ while in the second $j = |l - s|$. (This is illustrated in the case of $l = 1$ and $s = \frac{1}{2}$ in figure 6.2.) The other allowed values of j form a ladder with integer steps between these extremes.

In the case where s equals $\frac{1}{2}$ the possible values of j are $(l + \frac{1}{2})$ and $(l - \frac{1}{2})$—unless l equals zero when j and s are identical. Thus it follows from (6.47) that spin–orbit coupling results in all states with $l \neq 0$ being split into two, one member of the resulting doublet having its energy raised by an amount proportional to l while the other is lowered by an amount proportional to $(l + 1)$ (see figure 6.3). States with $l = 0$, however, have $j = s = \frac{1}{2}$ and are not

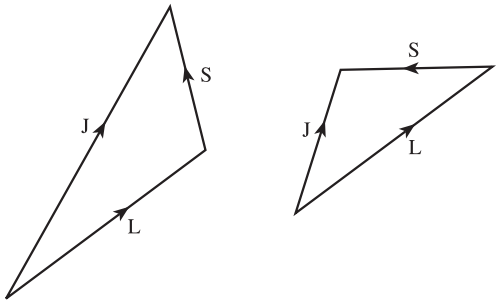


Figure 6.2. The maximum (minimum) possible values of **j** are obtained when **l** and **l** are as near parallel (antiparallel) as possible, given that their magnitudes must be proportional to $j(j+1)$, $l(l+1)$ and $s(s+1)$, where j, l and s are integers or half-integers. The diagram shows the case where $l = 1$, $s = \frac{1}{2}$ and $j = \frac{3}{2}$ or $\frac{1}{2}$.

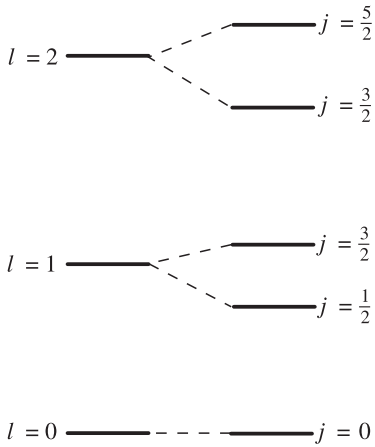


Figure 6.3. The effect of spin–orbit coupling on the energy levels of a one-electron atom.

split by spin–orbit coupling. We note that the quantum number m_j does not enter (6.47)—as indeed would be expected because the spherical symmetry of the problem has not been broken—so each state with a given value of j consists of $(2j+1)$ degenerate components which can be separated by an applied magnetic field as discussed in the next section.

A familiar example of such a doublet is observed in the spectrum of sodium and other alkali metals. The outer electron in sodium can be considered as moving in a spherically symmetric potential resulting from the nucleus and ten inner electrons. The well-known D lines result from transitions between states with $l = 1$ and $l = 0$ respectively. The first of these is split into two by the spin–orbit coupling, one component having $j = \frac{3}{2}$ and being fourfold degenerate and the

other having $j = \frac{1}{2}$ and being doubly degenerate, while the second is not affected by spin–orbit coupling and is therefore single. Transitions between the two sets of levels therefore produce a pair of closely spaced spectral lines in agreement with experimental observation. The difference between the frequencies of these lines can be calculated in the way described earlier and such calculations produce results that are in excellent agreement with experiment in every case.

Similar effects of spin–orbit coupling are observed in the spectrum of hydrogen, but these are complicated by the fact that states with different l are accidentally degenerate.

6.5.3 The weak-field Zeeman effect

The *weak-field Zeeman effect* refers to the case where the applied magnetic field B_0 is non-zero, but still so small that its contribution to the energy is much less than that due to spin–orbit coupling. In this case, we can assume the eigenfunctions to be effectively unchanged from the zero-field case—cf. the arguments leading to (6.46) and (6.47). A level with a particular value of j, l and s is split into $(2j + 1)$ states, each denoted by a particular value of the quantum number, m_j, associated with the z component of the total angular momentum. The evaluation of the magnitude of the splitting in the weak-field case is complicated by the fact that the operator representing the total magnetic moment $\hat{\boldsymbol{\mu}} = (e/2m_e)(\hat{\mathbf{L}} + 2\hat{\mathbf{S}})$ represents a vector which is not in the same direction as the operator representing the total angular momentum—$(\hat{\mathbf{L}} + \hat{\mathbf{S}})$ (see figure 6.4). We shall treat this problem more rigorously later, but for the moment use the following semi-classical argument. Referring to figure 6.4 , if $\mathbf{j} = \mathbf{l} + \mathbf{s}$, then for a given $\mathbf{j}$, $\mathbf{l}$ and $\mathbf{s}$, $\boldsymbol{\mu}$ must lie on a cone with axis $\mathbf{j}$. The average value of $\boldsymbol{\mu}$ is the average over the surface of the cone which is just the component (μ_j) of the total magnetic moment in the direction of the total angular momentum. We therefore have

$$\begin{aligned}\mu_j &= \mathbf{j}\cdot\boldsymbol{\mu}/j \\ &= \frac{e}{2m_e j}\mathbf{j}\cdot(\mathbf{l}+2\mathbf{s}) = \frac{e}{2m_e j}(j^2 + \mathbf{j}\cdot\mathbf{s}) \\ &= \frac{ej}{2m_e}\left(1 + \frac{\mathbf{l}\cdot\mathbf{s} + s^2}{j^2}\right) = \frac{ej}{2m_e}\left(1 + \frac{j^2 - l^2 + s^2}{2j^2}\right)\end{aligned}$$

using (6.41). The interaction energy δE is then given by $B\mu_{jz}$ where μ_{jz} is the z component of μ_j. That is

$$\delta E = \frac{eBj_z}{2m_e}\left(1 + \frac{j^2 - l^2 + s^2}{2j^2}\right) \tag{6.48}$$

We now replace the classical expressions for the magnitudes of the angular momenta by the eigenvalues of the corresponding quantum-mechanical operators,

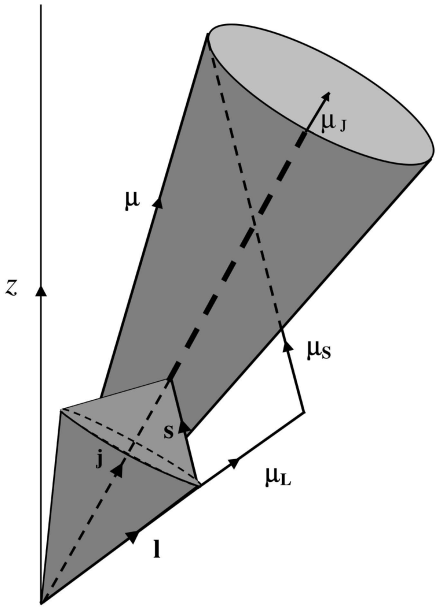


Figure 6.4. The orbital and spin magnetic moments are parallel to the respective angular momenta, but the constants of proportionality are different, so that the total magnetic moment, $\boldsymbol{\mu}$, is not parallel to **j**. The vectors representing **l**, **s**, **j**, $\boldsymbol{\mu}_l$, $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}$ are shown, assuming that they all lie in the plane of the page. However, for a given **j**, **l** and **s** lie on the surfaces of the cones and, as a result, $\boldsymbol{\mu}$ also lies on a cone whose axis is **j**. In weak magnetic fields, the average value of $\boldsymbol{\mu}$ is therefore its component parallel to **j**.

so that the change in energy relative to the zero-field case is given by

$$\delta E = g\mu_B m_j B \tag{6.49}$$

where $\mu_B = e\hbar/2m_e$ is the Bohr magneton (5.29) and

$$g = 1 + \frac{j(j+1) - l(l+1) + s(s+1)}{2j(j+1)} \tag{6.50}$$

is known as the Landé g-factor. We see, for example, that in the case where $s = 0$, j must equal l and therefore g equals 1. This corresponds to all the angular momentum being orbital so that the total magnetic moment is directly proportional to the total angular momentum. This also happens if there is no orbital angular momentum (that is, $l = 0$ and therefore $j = s$) and in this case $g = 2$.

6.6 A more general treatment of the coupling of angular momenta

We consider two angular momenta whose squared magnitudes are represented by the operators $\hat{L}^2$ and $\hat{S}^2$ and whose z components are represented by $\hat{L}_z$ and $\hat{S}_z$. Despite this notation, our treatment is not be confined to the spin–orbit case, but can be applied to any two coupled angular momenta. We assume that the two angular momenta can be measured compatibly so all four operators commute. That is,

$$\begin{aligned}[\hat{L}^2, \hat{L}_z] = [\hat{S}^2, \hat{S}_z] = [\hat{L}^2, \hat{S}^2] = [\hat{L}^2, \hat{S}_z] \\ = [\hat{L}_z, \hat{S}^2] = [\hat{L}_z, \hat{S}_z] = 0 \end{aligned} \tag{6.51}$$

The four operators therefore have a common set of eigenstates, which will be specified by the quantum numbers l, m_l, s and m_s, where these are all either integers or half-integers and where $-l \leqslant m_l \leqslant l$ and $-s \leqslant m_s \leqslant s$, as discussed in the previous chapter. We represent the state with these quantum numbers in Dirac notation as $|l, m_l, s, m_s\rangle$ or just $|m_l, m_s\rangle$. This state is therefore a direct product of the individual eigenstates. That is,

$$|m_l, m_s\rangle = |m_l\rangle|m_s\rangle \tag{6.52}$$

The fact that the expressions (6.52) are eigenfunctions of all four operators follows immediately on substitution into the appropriate eigenvalue equations.

This discussion has assumed that the values of the total and z component of both angular momenta are known. Although this condition is sometimes fulfilled at least approximately (as in the strong-field Zeeman effect) other situations, (such as spin–orbit coupling) often arise in which the values of the individual z components are unknown. The quantities measured in the latter case are, typically, the individual magnitudes whose squares are represented by $\hat{L}^2$ and $\hat{S}^2$ as before, along with the squared magnitude and z component of the total which are represented by the operators $\hat{J}^2$ and $\hat{J}_z$ respectively where

$$\hat{J}^2 = |\hat{\mathbf{L}} + \hat{\mathbf{S}}|^2 = \hat{L}^2 + \hat{S}^2 + 2\hat{\mathbf{L}} \cdot \hat{\mathbf{S}} \tag{6.53}$$

and

$$\hat{J}_z = \hat{L}_z + \hat{S}_z \tag{6.54}$$

Clearly $\hat{J}^2$ commutes with $\hat{L}^2$ and $\hat{S}^2$ and also with $\hat{J}_z$, but not with $\hat{L}_z$ or $\hat{S}_z$—these relations can be checked by substituting from (6.53) and (6.54) into the commutation relation (5.7). It follows from the general discussion of degeneracy at the end of chapter 4 that the operators $\hat{J}^2$, $\hat{J}_z$, $\hat{L}^2$ and $\hat{S}_z$ possess a common set of eigenstates. We now explain how to obtain these along with the corresponding eigenvalues.

The eigenvalues of the square of the magnitude of any angular momentum vector and its corresponding z component were determined in the previous

chapter, using a very general argument based only on the commutation relations between the operators representing the components. The results of this can therefore be applied to the present case, which means that the eigenvalues of $\hat{J}^2$, $\hat{J}_z$, $\hat{L}^2$, and $\hat{S}^2$ are $j(j+1)\hbar^2$, $m_j\hbar$, $l(l+1)\hbar^2$ and $s(s+1)\hbar^2$ respectively where j, l, s and m_j are integers or half-integers and $-j \leqslant m_j \leqslant j$. The common eigenstates can therefore be denoted by these four quantum numbers and represented by $|j, m_j, l, s\rangle$ or $|j, m_j\rangle$. To find expressions for these, we first consider again the products $|m_l\rangle|m_s\rangle$ which were shown earlier to be eigenfunctions of the operators $\hat{L}^2$, $\hat{S}^2$, $\hat{L}_z$ and $\hat{S}_z$. If we operate on these with $\hat{J}_z$ we obtain, using (6.54),

$$\begin{aligned}\hat{J}_z|m_l\rangle|m_s\rangle &= (\hat{L}_z + \hat{S}_z)|m_l\rangle|m_s\rangle \\ &= (m_l + m_s)\hbar|m_l\rangle|m_s\rangle \end{aligned} \tag{6.55}$$

It therefore follows that the products are also eigenfunctions of $\hat{J}_z$ with eigenvalues $m_j\hbar$ where

$$m_j = m_l + m_s \tag{6.56}$$

and that all such products that have the same values of l, s and m_j form a degenerate set with respect to the operators $\hat{L}^2$, $\hat{S}^2$ and $\hat{J}_z$, but not necessarily of $\hat{J}^2$. Physically this means that the total z component of angular momentum must be the sum of the two individual z components, and any allowed orientation of $\mathbf{L}$ and $\mathbf{S}$ that satisfies this condition and has the correct value for the total angular momentum is a possible state of the system. Remembering the general discussion of degenerate systems at the end of chapter 4, we see that any linear combination of these degenerate eigenfunctions is also an eigenfunction with the same eigenvalue, so we can express the eigenfunctions of $\hat{J}^2$ as linear combinations of the members of this degenerate set:

$$|j, m_j\rangle = \sum_{m_l=-l,l} \sum_{m_s=-s,s} C_{j,m_j,m_l,m_s}|m_l\rangle|m_s\rangle \tag{6.57}$$

We first consider the state where m_l and m_s have their maximum possible values which are l and s respectively. There is only one product, $|m_l = l\rangle|m_s = s\rangle$, corresponding to these quantum numbers which must therefore be the required eigenstate in this case. Moreover, it follows from (6.56) that the corresponding value of m_j is equal to $(l + s)$ and, as this is the maximum possible value of m_j, it must also be equal to j. We therefore have

$$|j = l + s, m_j = l + s\rangle = |m_l = l\rangle|m_s = s\rangle$$

or

$$|l + s, l + s\rangle = |l\rangle|s\rangle \tag{6.58}$$

We can now use the ladder operators introduced at the end of chapter 5 to obtain the state where $j = l + s$ and $m_j = j - 1$ from (6.58). We first

note that it follows directly from their definition (5.32) that the ladder operators corresponding to states of the combined angular momentum are just sums of these corresponding to the individual angular momenta. In particular

$$\hat{J}_- = \hat{L}_- + \hat{S}_- \tag{6.59}$$

Operating on (6.58) with $\hat{J}_-$ and using (6.59) we get

$$\hat{J}_-|l+s, l+s\rangle = |s\rangle\hat{L}_-|l\rangle + |l\rangle\hat{S}_-|s\rangle \tag{6.60}$$

Hence, remembering the normalization results (5.53) and (5.54)

$$\begin{aligned}\hbar^{-1}\big[(l+s)(l+s+1) - (l+s)(l+s-1)\big]^{1/2}&|l+s, l+s-1\rangle \\ &= |s\rangle\hbar^{-1}\big[l(l+1) - l(l-1)\big]^{1/2}|l-1\rangle \\ &\quad + |l\rangle\hbar^{-1}\big[s(s+1) - s(s-1)\big]^{1/2}|s-1\rangle\end{aligned} \tag{6.61}$$

i.e.

$$|l+s, l+s-1\rangle = \left(\frac{l}{l+s}\right)^{1/2}|l-1\rangle|s\rangle + \left(\frac{s}{l+s}\right)^{1/2}|l\rangle|s-1\rangle \tag{6.62}$$

There is a second state with $m_j = l+s-1$. It must have a different value of j and the only possibility is $l+s-1$. To be orthogonal to $|l+s, l+s-1\rangle$, it must have the form

$$|l+s-1, l+s-1\rangle = \left(\frac{s}{l+s}\right)^{1/2}|l-1\rangle|s\rangle - \left(\frac{l}{l+s}\right)^{1/2}|l\rangle|s-1\rangle \tag{6.63}$$

A further application of $\hat{J}_-$ to the states given in (6.62) and (6.63) produces expressions for the states $|l+s, l+s-2\rangle$ and $|l+s-1, l+s-2\rangle$; an expression for the state $|l+s-2, l+s-2\rangle$ is found using orthogonality—cf. (6.63). We note, however, that some of these states will be absent if l or s is less than 2. In general, the process can be continued by further applications of $\hat{J}_-$, which create states of smaller m_j, and adding in an additional value of j each time using orthogonality until the minimum value of j (i.e. $|l-s|$) is reached. Further applications of $\hat{J}_-$ reduce m_j further with values of j now being successively eliminated until m_j has its minimum value of $-j$. At this point there is again only one state : $|l+s, -l-s\rangle$. An entirely equivalent procedure is to start with this state and apply the raising operator $\hat{J}_+$ until $m_j = j$.

Table 6.2 shows the results of applying this procedure to the particular case where $l = 1$ and $s = \frac{1}{2}$. The second row is obtained by substituting into (6.62) and the third and fourth by further applications of $\hat{J}_-$. The fifth row is obtained by substituting into (6.63), while the sixth is obtained from the fifth by applying $\hat{J}_-$.

Table 6.2. Clebsch–Gordan coefficients for the case $l = 1$ and $s = \frac{1}{2}$. Eigenfunctions of the total angular momentum with given values of (j, m_j) are constructed as linear combinations of products of the angular momentum eigenfunctions denoted by the values of (m_l, m_s).

	(m_l, m_s)					
(j, m_j)	$(1, \frac{1}{2})$	$(1, -\frac{1}{2})$	$(0, \frac{1}{2})$	$(0, -\frac{1}{2})$	$(-1, \frac{1}{2})$	$(-1, -\frac{1}{2})$
$(\frac{3}{2}, \frac{3}{2})$	1	0	0	0	0	0
$(\frac{3}{2}, \frac{1}{2})$	0	$\frac{1}{\sqrt{3}}$	$\frac{2}{\sqrt{3}}$	0	0	0
$(\frac{3}{2}, -\frac{1}{2})$	0	0	0	$\frac{2}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	0
$(\frac{3}{2}, -\frac{3}{2})$	0	0	0	0	0	1
$(\frac{1}{2}, \frac{1}{2})$	0	$\sqrt{\frac{2}{3}}$	$-\frac{1}{\sqrt{3}}$	0	0	0
$(\frac{1}{2}, -\frac{1}{2})$	0	0	0	$-\frac{1}{\sqrt{3}}$	$\sqrt{\frac{2}{3}}$	0

The process gets very complex for large quantum numbers, and it is easier to look up the results in a reference text or via the internet. Once the Clebsch–Gordan coefficients have been determined, they can be used to construct a matrix representation of the energy operator from which the energy eigenvalues can be obtained. We will describe this for the case of the Hamiltonian operator given in (6.43) which we re-write as

$$\begin{aligned}\hat{H}' &\equiv \hat{H}'^{(1)} + \hat{H}'^{(2)} \\ &= \langle f(r) \rangle (\hat{J}^2 - \hat{L}^2 - \hat{S}^2) - \frac{eB_0}{2m_e}(\hat{L}_z + 2\hat{S}_z)\end{aligned} \tag{6.64}$$

the expectation value being taken over the radial coordinate only. Using the states $|j, m_j\rangle$ as a basis, we calculate the matrix elements

$$\begin{aligned}H'_{(j,m_j),(j',m'_j)} &= H'^{(1)}_{(j,m_j),(j',m'_j)} + H'^{(2)}_{(j,m_j),(j',m'_j)} \\ &= \langle f(r) \rangle \langle j, m_j | (\hat{J}^2 - \hat{L}^2 - \hat{S}^2) | j', m'_j \rangle \\ &\quad - \frac{eB_0}{2m_e} \langle j, m_j | (\hat{L}_z + 2\hat{S}_z) | j', m'_j \rangle\end{aligned} \tag{6.65}$$

$|j, m_j\rangle$ is an eigenstate of $(\hat{J}^2 - \hat{L}^2 - \hat{S}^2)$ with eigenvalue $[j(j+1) - l(l+1) - s(s+1)]\hbar^2$, so it follows (remembering orthogonality) that the first term in (6.65) contributes only to the diagonal elements of the matrix and we have

$$H'^{(1)}_{(j,m_j),(j',m'_j)} = \langle f(r) \rangle [j(j+1) - l(l+1) - s(s+1)] \delta_{j,j'} \delta_{m_j,m'_j} \tag{6.66}$$

Table 6.3. The matrix representing $\hat{H}'$ in the case where $l = 1$ and $s = \frac{1}{2}$. $\epsilon = \langle f(r)\rangle\hbar^2$ and $\mu_B = e\hbar/2m_e$.

	(j, m_j)					
(j, m_j)	$(\frac{3}{2}, \frac{3}{2})$	$(\frac{3}{2}, -\frac{1}{2})$	$(\frac{3}{2}, -\frac{1}{2})$	$(\frac{3}{2}, -\frac{3}{2})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, -\frac{1}{2})$
$(\frac{3}{2}, \frac{3}{2})$	$\epsilon - 2\mu_B B$	0	0	0	0	0
$(\frac{3}{2}, \frac{1}{2})$	0	$\epsilon - \frac{2}{3}\mu_B B$	0	0	$\frac{\sqrt{2}}{3}\mu_B B$	0
$(\frac{3}{2}, -\frac{1}{2})$	0	0	$\epsilon + \frac{2}{3}\mu_B B$	0	0	$\frac{\sqrt{2}}{3}\mu_B B$
$(\frac{3}{2}, -\frac{3}{2})$	0	0	0	$\epsilon + 2\mu_B B$	0	0
$(\frac{1}{2}, \frac{1}{2})$	0	$\frac{\sqrt{2}}{3}\mu_B B$	0	0	$-2\epsilon - \frac{1}{3}\mu_B B$	0
$(\frac{1}{2}, -\frac{1}{2})$	0	0	$\frac{\sqrt{2}}{3}\mu_B B$	0	0	$-2\epsilon + \frac{1}{3}\mu_B B$

We evaluate the second term by expressing the states $|j, m_j\rangle$ in terms of the products $|m_l\rangle|m_s\rangle$ using Clebsch–Gordan coefficients as described earlier, and get

$$\begin{aligned} H'^{(2)}_{(j,m_j),(j',m'_j)} &= -\frac{eB}{2m_e}\sum_{m_l,m_s} C^*_{(j,m_j)(m_l,m_s)}\langle m_s|\langle m_l|(\hat{L}_z + 2\hat{S}_z) \\ &\quad \times \sum_{m'_l,m'_s} C_{(j',m'_j)(m'_l,m'_s)}|m'_l\rangle|m'_s\rangle \\ &= -\frac{eB}{2m_e}\sum_{(m_l,m_s)(m'_l,m'_s)} C^*_{(j,m_j)(m_l,m_s)}C_{(j',m'_j)(m'_l,m'_s)} \\ &\quad \times \langle m_l|\langle m_s|(\hat{L}_z + 2\hat{S}_z)|m'_l\rangle|m'_s\rangle \end{aligned} \tag{6.67}$$

We can now use the fact that $|m_l\rangle$ and $|m_s\rangle$ are eigenstates of $\hat{L}_z$ and $\hat{S}_z$ respectively, along with orthogonality to get

$$H'^{(2)}_{(j,m_j),(j',m'_j)} = -\frac{e\hbar B}{2m_e}\sum_{m_l,m_s} C^*_{(j,m_j)(m_l,m_s)}C_{(j',m'_j)(m_l,m_s)}(m_l + 2m_s) \tag{6.68}$$

(6.66) and (6.68) can be combined to form the complete Hamiltonian matrix, which can then be diagonalized to obtain the energy eigenvalues. Following this procedure and using the Clebsch–Gordan coefficients in table 6.2, we get a matrix whose elements are listed in table 6.3 for the case where $l = 1$ and $s = \frac{1}{2}$.

We can check that this gives the same answers as those we obtained in the limits of large and small field discussed earlier. Considering the zero-field case first, we see that the matrix in table 6.3 becomes diagonal when $B = 0$; the four

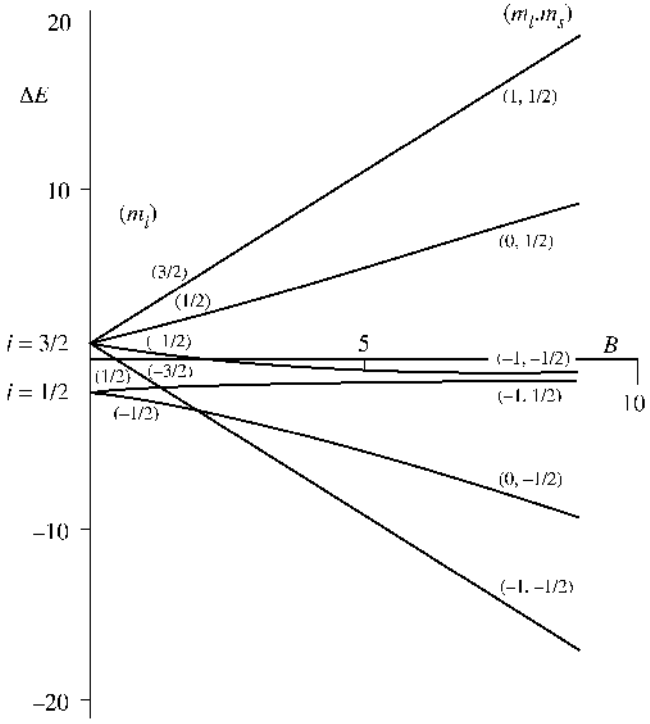


Figure 6.5. The splitting ΔE of the states with $l = 1$ in a one-electron atom as a function of applied magnetic field. Energy is in units of ϵ and the field is in units of ϵ/μ_B.

states with $j = \frac{3}{2}$ all have spin–orbit energies equal to ϵ, while the two with $j = \frac{1}{2}$ have spin–orbit energies of -2ϵ, agreeing with (6.47) and figure 6.3.

The strong-field limit corresponds to $\epsilon = 0$ in table 6.3; the matrix now factorizes into two 1×1 matrices where the Zeeman energies are $\pm2\mu_B B$, and two 2×2 matrices whose eigenvalues can be straightforwardly shown to be $\pm\mu_B B$ and zero. These results are exactly what we obtained earlier—(6.45). One advantage of the general formulation is that the calculation can be made for any values of ϵ and μB, and figure 6.5 shows the results of an evaluation of the eigenvalues of $\hat{H}'$ for a wide range of fields.

Although we have considered only a few comparatively simple examples, the principles underlying the addition of angular momenta have wide application. For example, similar techniques can be applied to the problem of the coupling of the orbital and spin angular momenta of many-electron atoms when detailed calculations show that this can often be considered as a two-stage process: first the orbital and spin angular momenta of the different electrons separately combine to form total orbital and spin vectors; then these two quantities interact in a similar

way to that described earlier in the case of one-electron atoms. Such a system is described as Russell–Saunders or L–S coupling, and provides a satisfactory account of the atomic spectra of atoms with low atomic number. Heavy elements, however, are better described by 'j–j coupling' in which the orbital and spin angular momenta of each electron interact strongly to form a total described by the quantum number j. The totals associated with different electrons then combine together to form a grand total for the atom.

The application of many of the ideas discussed in this chapter and the last to the fields of nuclear and particle physics are also of great importance. This is not only due to the importance of angular momentum in these systems, but also because other physical quantities turn out to be capable of representation by a set of operators obeying the same algebra as do those representing angular momentum components. An example of this is the quantity known as *isotopic spin*. Using this concept the proton and neutron can be described as separate states of the same particle, assuming that the total isotopic-spin quantum number is $\frac{1}{2}$ so that its 'z component' can be represented by a quantum number having the values $\frac{1}{2}$ and $-\frac{1}{2}$. Properties such as the charge and mass differences between the two states can then be treated as resulting from isotopic-spin-dependent interactions between the three quarks that constitute the nucleon. It is also found that excited states of the nucleon whose properties can be explained by assigning appropriate values to the two isotopic-spin quantum numbers exist.

Problems

6.1 Show that the Hermitian matrices

$$[x] = \left(\frac{1}{2}\frac{\hbar}{m\omega}\right)^{1/2} \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & \sqrt{2} & 0 & \cdots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \cdots \\ 0 & 0 & \sqrt{3} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}$$

and

$$[P] = \left(\frac{1}{2}m\omega\hbar\right)^{1/2} \begin{bmatrix} 0 & -i & 0 & 0 & \cdots \\ i & 0 & -i/\sqrt{2} & 0 & \cdots \\ 0 & i\sqrt{2} & 0 & -i\sqrt{3} & \cdots \\ 0 & 0 & i\sqrt{3} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}$$

obey the correct commutation rules for position and momentum in one dimension. Show that these can be used to obtain the energy levels of a harmonic oscillator of classical frequency ω and compare your answers with those obtained in chapter 2.

6.2 Calculate the expectation values $\langle \hat{S}_x \rangle$, $\langle \hat{S}_y \rangle$, $\langle \hat{S}_x^2 \rangle$, and $\langle \hat{S}_y^2 \rangle$ for a spin-half particle known to be in an eigenstate of $\hat{S}_z$. Show that the product $\langle \hat{S}_x^2 \rangle \langle \hat{S}_y^2 \rangle$ is consistent with the uncertainty principle.

6.3 Obtain matrices representing the raising and lowering operators in the case of a spin-half system. Verify that these have the general properties attributed to ladder operators in chapter 5.

6.4 A spin-half particle, initially in an eigenstate of $\hat{S}_x$ with eigenvalue $\frac{1}{2}\hbar$, enters a Stern–Gerlach apparatus oriented to measure its angular momentum component in a direction whose orientation with

respect to the x and z axes is defined by the spherical polar angles θ and ϕ. Obtain expressions for the probabilities of obtaining the results $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ from the second measurement.

6.5 Spin-half particles, initially in an eigenstate of $\hat{S}_z$ with eigenvalue $\frac{1}{2}\hbar$, are directed into a Stern–Gerlach apparatus oriented to measure the x component of spin. The lengths of the two possible paths through the second apparatus are precisely equal and the particles are then directed into a common path so that it is impossible to tell which route any particular particle followed. What result would you expect to obtain if the z component of the particle spin is now measured? Discuss this from the point of view of the quantum theory of measurement and compare your discussion with that of the two-slit experiment discussed in chapter 1.

6.6 Show that the following matrices obey the appropriate commutation rules and have the correct eigenvalues to represent the three components of angular momentum of a spin-one particle:

$$[L_x] = \frac{\hbar}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad [L_y] = \frac{\hbar}{\sqrt{2}}\begin{bmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{bmatrix} \qquad [L_z] = \hbar\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Verify that the corresponding matrix representing the square of the total angular momentum also has the correct eigenvalues.

6.7 Obtain the eigenvectors of the matrices given in problem 6.6 and use these to find the relative probabilities of the possible results of the measurement of the x component of spin on a spin-one particle initially in an eigenstate of S_z with eigenvalue $\hbar$.

6.8 Discuss the splitting of the $l = 2$ state of a one-electron atom due to (i) spin–orbit coupling, (ii) the strong-field Zeeman effect and (iii) the weak-field Zeeman effect.

Chapter 7

Time-independent perturbation theory and the variational principle

Throughout our discussion of the general principles of quantum mechanics we have emphasized the importance of the eigenvalues and eigenfunctions of the operators representing physical quantities. However, we have seen that solving the basic eigenvalue equations determining these is often not straightforward. In the early chapters discussing the energy eigenvalue equation (or time-independent Schrödinger equation) for example, we found that this could often not be solved exactly and that, even when a solution was possible, it frequently required considerable mathematical analysis. Because of this, methods have been developed to obtain approximate solutions to eigenvalue equations. One of the most important of such techniques is known as time-independent perturbation theory and will be the first method to be discussed in the present chapter. Another, known as the variational principle, will be discussed later. We shall confine our discussion to the particular case of the energy eigenvalue equation and use the wavefunction representation, but we shall re-state some of the important results in Dirac notation so that they can be readily transformed to a matrix equation if desired.

Perturbation theory can be applied when the Hamiltonian operator $\hat{H}$, representing the total energy of the system, can be written in the form

$$\hat{H} = \hat{H}_0 + \hat{H}' \tag{7.1}$$

where the eigenvalues E_{0n} and eigenfunctions u_{0n} of $\hat{H}_0$ are assumed to be known and the operator $\hat{H}'$ represents an additional energy known as a *perturbation*, which is in some sense small compared with $\hat{H}_0$. Thus if we know the solution to a problem described by $\hat{H}_0$ (for example the energy eigenvalues and eigenfunctions of the hydrogen atom) we can use perturbation theory to obtain approximate solutions to a related problem (such as the energy eigenvalues and eigenfunctions of a hydrogen atom subject to a weak electric field). We shall use (7.1) along with

the eigenvalue equation for the unperturbed system

$$\hat{H}_0 u_{0n} = E_{0n} u_{0n} \tag{7.2}$$

to express the eigenvalues and eigenfunctions of $\hat{H}$ in the form of a series, the leading (zero-order) term of which is independent of $\hat{H}'$ while the next (first-order) term contains expressions linear in $\hat{H}'$ and so on; a similar series is also obtained for the energy eigenfunctions. The mathematical complexity of the series increases rapidly with ascending order, and perturbation theory is generally useful only if the series converges rapidly—which usually means that the additional energy due to the perturbation is much smaller than the energy difference between typical neighbouring levels. Accordingly, we shall confine our treatment to obtaining expressions for the eigenvalues of $\hat{H}$ which are correct to second order in $\hat{H}'$ as well as first-order corrections to the eigenfunctions. We shall illustrate the perturbation method by considering its application to several physical problems.

7.1 Perturbation theory for non-degenerate energy levels

We first consider the effect of perturbations on energy levels that are not degenerate and return to the degenerate case later. We assume that the correct eigenvalues and eigenfunctions of $\hat{H}$ can be expressed as a series whose terms are of zeroth, first, second, etc., order in the perturbation $\hat{H}'$, and derive expressions for each of these in turn. This can be done most conveniently if we rewrite (7.1) in the form

$$\hat{H} = \hat{H}_0 + \beta \hat{H}' \tag{7.3}$$

where β is a constant, and obtain results that are valid for all values of β, including the original case where $\beta = 1$. It is important to note that β is not assumed to be small, but first, second etc. order corrections are identified as terms in β, β^2, etc. The series for E_n and u_n are written as

$$\left.\begin{aligned} E_n &= E_{0n} + \beta E_{1n} + \beta^2 E_{2n} + \cdots \\ u_n &= u_{0n} + \beta u_{1n} + \beta^2 u_{2n} + \cdots \end{aligned}\right\} \tag{7.4}$$

where the terms independent of β are known as zeroth-order terms, those in β are first-order, those in β^2 second-order, and so on. We substitute these expressions into the energy eigenvalue equation

$$\hat{H} u_n = E_n u_n \tag{7.5}$$

and obtain

$$\begin{aligned} &(\hat{H}_0 + \beta \hat{H}')(u_{0n} + \beta u_{1n} + \beta^2 u_{2n} + \cdots) \\ &\quad = (E_{0n} + \beta E_{1n} + \beta^2 E_{2n} + \cdots)(u_{0n} + \beta u_{1n} + \beta^2 u_{2n} + \cdots) \end{aligned} \tag{7.6}$$

Expanding this equation and equating the coefficients of the different powers of β we get

$$\hat{H}_0 u_{0n} = E_{0n} u_{0n} \tag{7.7}$$

$$\hat{H}' u_{0n} + \hat{H}_0 u_{1n} = E_{0n} u_{1n} + E_{1n} u_{0n} \tag{7.8}$$

$$\hat{H}' u_{1n} + \hat{H}_0 u_{2n} = E_{0n} u_{2n} + E_{1n} u_{1n} + E_{2n} u_{0n} \tag{7.9}$$

We note first that (7.7) is identical to (7.2) which is to be expected as the former refers to a perturbation of zero order and the latter describes the unperturbed system. We can obtain expressions for the first-order corrections from (7.8) by expressing u_{1n} as a linear combination of the complete set of unperturbed eigenfunctions u_{0k}:

$$u_{1n} = \sum_k a_{nk} u_{0k} \tag{7.10}$$

Substituting (7.10) into (7.8) gives, after rearranging and using (7.7),

$$(\hat{H}' - E_{1n}) u_{0n} = \sum_k a_{nk} (E_{0n} - E_{0k}) u_{0k} \tag{7.11}$$

We now multiply both sides of (7.11) by u_{0n}^* and integrate over all space, using the fact that the u_{0k} are orthonormal to get

$$E_{1n} = H'_{nn} \tag{7.12}$$

where

$$H'_{nn} = \int u_{0n}^* \hat{H}' u_{0n} \, d\tau \equiv \langle 0n | \hat{H}' | 0n \rangle \tag{7.13}$$

We have therefore obtained an expression for the first-order correction to the energy eigenvalues of the perturbed system in terms of the perturbation operator H' and the eigenfunctions of the unperturbed system, which are assumed to be known. We note that this first-order correction to the energy is just the expectation value of the perturbation operator, calculated using the unperturbed eigenfunctions. We can now obtain an expression for the first-order correction to the eigenfunction by multiplying both sides of (7.11) by u_{0m}^* (where $m \neq n$) and again integrate over all space when we get (again using orthonormality)

$$a_{nm} = \frac{H'_{mn}}{E_{0n} - E_{0m}} \qquad m \neq n \tag{7.14}$$

where H'_{mn} is defined by a similar equation to (7.13) with u_{0n}^* replaced by u_{0m}^* and is therefore a *matrix element* (cf. chapter 6). Hence, using (7.4) and (7.10),

$$u_n = (1 + a_{nn}) u_{0n} + \sum_{k \neq n} \frac{H'_{kn}}{E_{0n} - E_{0k}} u_{0k} + \text{higher order terms} \tag{7.15}$$

We shall now show that a_{nn} can be put equal to zero. To do this we note that (7.15) must be normalized. That is

$$1 = \int u_n^* u_n \, d\tau$$
$$= \int \left[(1 + a_{nn}^*) u_{0n}^* + \sum_{k \neq n} a_{nk}^* u_{0k}^* \right] \left[(1 + a_{nn}) u_{0n} + \sum_{k \neq n} a_{nk} u_{0k} \right] d\tau$$

Remembering again that the u_{0k} are orthonormal and retaining only first-order terms we get

$$1 + a_{nn}^* + a_{nn} = 1$$

Hence

$$a_{nn} = -a_{nn}^* \tag{7.16}$$

and a_{nn} is therefore an imaginary number which we can write as $i\gamma_n$ where γ_n is real. Hence

$$1 + a_{nn} = 1 + i\gamma_n \simeq \exp(i\gamma_n)$$

where the final equality is correct to first order as is (7.15). It follows that the factor $(1 + a_{nn})$ in the first term of (7.15) is equal to $\exp(i\gamma_n)$ to first order and the effect of this is simply to multiply u_{0n} by a phase factor. Referring to (7.13) we see that the first-order change in the energy eigenvalue is independent of the phase of u_{0n} and from (7.15) we see that the value of γ_n affects only the overall phase of the eigenfunction—at least as far as zero- and first-order terms are concerned. But we know that the absolute phase of an eigenfunction is arbitrary, so there is no loss of generality involved if we put $\gamma_n = 0$, leading to the following expression for the eigenfunction which is correct to first order:

$$u_n = u_{0n} + \sum_{k \neq n} \frac{H'_{kn}}{E_{0n} - E_{0k}} u_{0k} \tag{7.17}$$

or, in Dirac notation,

$$|n\rangle = |0n\rangle + \sum_{k \neq n} \frac{\langle 0n|\hat{H}'|0n\rangle}{E_{0n} - E_{0k}} |0k\rangle + \text{higher order terms} \tag{7.18}$$

Proceeding now to consider second-order terms, we first expand u_{2n} in terms of the unperturbed eigenfunctions

$$u_{2n} = \sum_k b_{nk} u_{0k} \tag{7.19}$$

and then substitute from (7.7), (7.10) and (7.19) into (7.9) which gives, after some rearrangement,

$$\sum_k b_{nk}(E_{0k} - E_{0n}) u_{0k} + \sum_k a_{nk}(\hat{H}' - E_{1n}) u_{0k} = E_{2n} u_{0n} \tag{7.20}$$

Multiplying (7.20) by u_{0n}^* and integrating over all space leads to

$$\begin{aligned} E_{2n} &= \sum_{k\neq n} a_{nk} H'_{nk} \\ &= \sum_{k\neq n} \frac{H'_{kn} H'_{nk}}{E_{0n} - E_{0k}} \\ &= \sum_{k\neq n} \frac{|H'_{kn}|^2}{E_{0n} - E_{0k}} \end{aligned} \tag{7.21}$$

where we have used (7.14) and the relation $H'_{kn} = H'^*_{nk}$, which follows from the definition of the matrix element and the fact that $\hat{H}'$ is a Hermitian operator. This procedure can be continued to obtain expressions for the second-order change in the eigenfunction as well as higher-order corrections to both the eigenfunctions and the eigenvalues. However, the expressions rapidly become more complicated and if the perturbation series does not converge rapidly enough for (7.12), (7.17) and (7.21) to be sufficient, it is usually better to look for some other method of solving the problem.

Example 7.1 The anharmonic oscillator We consider the case of a particle of mass m subject to a one-dimensional potential $V(x)$ where

$$V = \tfrac{1}{2} m\omega^2 x^2 + \gamma x^4 \tag{7.22}$$

and we wish to calculate the energy of the ground state to first order in γ. If γ were zero, the potential would correspond to a harmonic oscillator of classical frequency ω whose energy levels were shown in chapter 2 to be

$$E_{0n} = (n + \tfrac{1}{2})\hbar\omega \tag{7.23}$$

The corresponding ground-state energy eigenfunction was also evaluated in chapter 2 as

$$u_{00} = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega}{2\hbar} x^2\right) \tag{7.24}$$

The inclusion of the term γx^4 in the potential (7.22) changes the problem from a harmonic oscillator to an anharmonic oscillator. If γ is small, we can apply perturbation theory and the first-order correction to the ground-state energy is obtained by substituting from (7.24) into (7.12) giving

$$\begin{aligned} E_{10} &= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{-\infty}^{\infty} \gamma x^4 \exp\left(-\frac{m\omega}{\hbar} x^2\right) dx \\ &= \frac{3\hbar^2}{4m^2\omega^2}\gamma \end{aligned} \tag{7.25}$$

Table 7.1. The ground-state energy of an anharmonic oscillator calculated using first-order perturbation theory (E_{01}) and by numerical methods (E_0).

$g = \dfrac{2\hbar\gamma}{m^2\omega^3}$	$E_{01}/(\frac{1}{2}\hbar\omega)$	$E_0/(\frac{1}{2}\hbar\omega)$
0.01	1.007 50	1.007 35
0.1	1.075	1.065
0.2	1.15	1.12

which, along with the zero-order term $\frac{1}{2}\hbar\omega$, constitutes the required expression for the ground-state energy. If we characterize the 'strength' of the perturbation by the dimensionless quantity g defined as $g = (2\hbar/m^2\omega^3)\gamma$, it follows from (7.23) and (7.25) that the ratio of the first-order correction to the unperturbed energy is just $3g/4$. The perturbed energies calculated in this way are shown in table 7.1 for several values of g, along with values of the total energy (as a fraction of $\frac{1}{2}\hbar\omega$) calculated by solving the Schrödinger equation numerically with the potential (7.22). We see that when the total correction to the zero-order energy is around 1%, the error involved in using first-order perturbation theory is less than 0.02% of the total, and that the approximation is still correct to about 3% when the total correction is about 20%.

Example 7.2 The atomic polarizability of hydrogen The d.c. polarizability of an atom α is defined by the equation $\boldsymbol{\mu} = \alpha\varepsilon_0\mathcal{E}$ where $\boldsymbol{\mu}$ is the electric dipole moment induced by a steady uniform electric field $\mathcal{E}$. We shall consider a hydrogen atom in its ground state for which the wavefunction is spherically symmetric so that the direction of the z axis can be chosen as parallel to the field. A field of magnitude $\mathcal{E}$ in this direction will contribute an extra term, $\hat{H}'$, to the Hamiltonian which can be treated as a perturbation. Thus from elementary electrostatics

$$\hat{H}' = e\mathcal{E}z \tag{7.26}$$

where z is the coordinate of the electron with respect to the proton at the origin. Substituting from (7.26) into (7.17), the ground-state eigenfunction u_0 is given by

$$u_0 = u_{00} + e\mathcal{E}\sum_{k\neq 0}\frac{z_{k0}}{E_{00} - E_{0k}}u_{0k} \tag{7.27}$$

Classically, if the electron were at a position $\mathbf{r}$ with respect to the nucleus, the atom would have a dipole moment $\boldsymbol{\mu}$ where $\boldsymbol{\mu} = -e\mathbf{r}$. It follows that the quantum-mechanical operator representing the dipole moment has the same form and the expectation value of its z component is therefore $\langle\mu\rangle$ where

$$\langle\mu\rangle = -e\int u_0^* z u_0\, d\tau$$

and the components in the x and y directions clearly vanish from symmetry considerations. Using (7.27) we have

$$\langle \mu \rangle = -e \int u_{00}^* z u_{00}\, d\tau + 2e^2 \mathcal{E} \sum_{k \neq 0} \frac{|z_{k0}|^2}{E_{0k} - E_{00}} \tag{7.28}$$

to first order in $\mathcal{E}$, remembering that $z_{k0}^* = z_{0k}$. The first term on the right-hand side of (7.28) vanishes because u_{00} is spherically symmetric and z is an odd function, so it follows that the atomic polarizability, α, is given by

$$\alpha = \frac{2e^2}{\varepsilon_0} \sum_{k \neq 0} \frac{|z_{k0}|^2}{E_{0k} - E_{00}} \tag{7.29}$$

An alternative derivation which yields an identical expression to (7.29) consists of considering the perturbation expression for the energy: the first-order contribution (7.13) is proportional to z_{00} and is therefore zero on symmetry grounds, and the second-order term (7.21) is proportional to $\mathcal{E}^2$. The latter is then compared with the energy of an induced dipole in the applied field which equals $-\frac{1}{2}\alpha\varepsilon_0\mathcal{E}^2$.

The summation in (7.29) is over all the excited states of the hydrogen atom (including the continuum of unbound states) and considerable computational effort would be required to evaluate it exactly. However, we can calculate a maximum possible value or 'upper bound' for α using what is known as the Unsöld closure principle. In this case we replace each energy difference $E_{0k} - E_{00}$ in (7.29) by its smallest possible value, which is clearly the difference between the ground and first excited energy levels. Writing this quantity as ΔE we get

$$\begin{aligned} \varepsilon_0 \alpha &\leqslant \frac{2e^2}{\Delta E} \sum_{k \neq 0} |z_{k0}|^2 \\ &= \frac{2e^2}{\Delta E} \left(\sum_k z_{0k} z_{k0} - |z_{00}|^2 \right) \end{aligned} \tag{7.30}$$

As previously mentioned, the second term in (7.30) vanishes for symmetry reasons. To evaluate the first term we refer back to the discussion of matrix representations in chapter 6, from which it follows that the quantities z_{0k} constitute the elements of a matrix representing the operator z. If we now apply the standard rules for multiplying matrices we see that the first term in (7.30) is equal to the leading diagonal element of a similar matrix representing the operator

z^2. Hence

$$\sum_k z_{0k} z_{k0} = (z^2)_{00} = \int u_{00}^* z^2 u_{00}\, d\tau$$
$$= \frac{1}{\pi a_0^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \exp\left(-\frac{2r}{a_0}\right) r^4 \cos^2\theta \sin\theta\, dr\, d\theta\, d\phi$$
$$= a_0^2 \tag{7.31}$$

where we have used spherical polar coordinates and the standard expression for the ground-state eigenfunction (cf. (3.70)).

The energy difference ΔE is obtained from the expressions (3.72) for the energy levels of the hydrogen atom

$$\Delta E = \frac{3e^2}{8(4\pi\varepsilon_0)a_0} \tag{7.32}$$

Substituting from (7.31) and (7.32) into (7.30) we get

$$\alpha \leqslant 64\pi a_0^3/3$$
$$= 67.02 a_0^3 \tag{7.33}$$

This maximum value is within 15% of the experimental value of $57.8a_0^3$. Because the upper bound can be evaluated using only the ground-state eigenfunction and the difference between the energies of the two lowest states, this approximation can be usefully applied to the calculation of polarizabilities of other systems, even where the full set of matrix elements and energy differences are not known and expressions such as (7.29) cannot be evaluated. Moreover, as we shall see later in this chapter, the variational principle can often be used to generate a lower bound to α so that by combining both methods, a theoretical estimate can be made with a precision that is rigorously known.

7.2 Perturbation theory for degenerate levels

The application of perturbation theory to degenerate systems can often lead to powerful insights into the physics of the systems considered. This is because the perturbation often 'lifts' the degeneracy of the energy levels, leading to additional structure in the line spectrum. We saw examples of this in chapter 6 where the inclusion of a spin–orbit coupling term split the previously degenerate energy levels (figure 6.3) and further splitting resulted from the application of magnetic fields (figure 6.5). We shall now develop a general method for extending perturbation theory to such degenerate systems.

The mathematical reason why we cannot apply the theory in the form developed so far to the degenerate case follows from the fact that if one or

more of the energy levels E_{0k} in (7.15) is equal to E_{0n} then at least one of the denominators $(E_{0n} - E_{0k})$ in (7.15) would be zero, leading to an infinite value for the corresponding term in the series. To understand the reason why this infinity arises, we consider the case of twofold degeneracy and take u_{01} and u_{02} to be two eigenfunctions of the unperturbed Hamiltonian $\hat{H}_0$ with eigenvalue E_{01}. It is important to remember (see the discussion of degeneracy in chapter 4) that any linear combination of u_{01} and u_{02} is also an eigenfunction of H_0 with the same eigenvalue, so that if we had solved the unperturbed problem in a different way we could well have come up with a different pair of eigenfunctions. Suppose that a small perturbation $\hat{H}'$ is applied and that as a result we have two states of slightly different energy whose eigenfunctions are v_1 and v_2; because the system is no longer degenerate, linear combinations of these are not energy eigenfunctions. Now imagine we remove the perturbation gradually. As it tends to zero, v_1 and v_2 will tend to eigenfunctions of the unperturbed system; let these be u'_1 and u'_2. In general u'_1 and u'_2 will be different from u_1 and u_2, although the former quantities will of course be linear combinations of the latter. Hence, if we start from u_1 and u_2 the application of a very small perturbation has to produce large changes in the eigenfunctions to get us to v_1 and v_2. This is why the infinite terms appear in the perturbation series.

The way out of this apparent impasse is to use other methods to obtain the correct starting functions u'_1 and u'_2. Referring back to (7.11) we see that, if u_{01} is one of a degenerate pair whose other member is u_{02}, then the coefficient a_{12} on the right-hand side is indeterminate (because $E_{01} = E_{02}$) and the infinity arises later in the derivation when we divide through by $E_{01} - E_{02}$ which equals zero. However, the correct zero-order eigenfunction must be some linear combination of u_{01} and u_{02} which we can write as

$$v_0 = C_1 u_{01} + C_2 u_{02} \tag{7.34}$$

where the constants C_1 and C_2 are to be determined. Using this in place of u_{01}, the equivalent equation to (7.11) becomes

$$(\hat{H}' - E_1)(C_1 u_{01} + C_2 u_{02}) = \sum_k a_{1k}(E_{01} - E_{0k})u_{0k} \tag{7.35}$$

where the first-order correction to the energy of the degenerate states is now written as E_1. We now multiply both sides of (7.35) by u^*_{01} and integrate over all space, then repeat the process using u^*_{02} to obtain the following pair of equations:

$$\left.\begin{aligned}(H'_{11} - E_1)C_1 + H'_{12}C_2 = 0\\ H'_{21}C_1 + (H'_{22} - E_1)C_2 = 0\end{aligned}\right\} \tag{7.36}$$

where we have assumed that u_{01} and u_{02} are orthogonal because it was shown in chapter 4 that, although degenerate eigenfunctions need not be orthogonal, an orthogonal set can always be generated using Schmidt orthogonalization.

Equations (7.36) have a non-trivial solution if, and only if, the determinant of the coefficients is zero:

$$\begin{vmatrix} H'_{11} - E_1 & H'_{12} \\ H'_{21} & H'_{22} - E_1 \end{vmatrix} = 0 \tag{7.37}$$

The solution of (7.37) leads to two possible values for E_1. Substituting these into (7.36), and using the condition that the zero-order eigenfunction must be normalized, leads to two sets of values for the coefficients C_1 and C_2. Thus we have found two zero-order functions, v_{01} and v_{02} that are linear combinations of u_{01} and u_{02} and which, in general, correspond to different first-order corrections to the energy. Hence, as we expected, one effect of the perturbation may be to remove the degeneracy. These two linear combinations represent the appropriate choice of zero-order eigenfunctions for the problem and they can be used, along with the unperturbed eigenfunctions for the other states, to evaluate higher-order corrections in a similar way to the non-degenerate case. Note that diagonalizing the matrix in (7.37) means that the matrix element $\int v_{01}^* \hat{H}' v_{02} = 0$, and the infinite terms in the expansion have been removed.

This discussion has been confined to the case of twofold degeneracy, but the extension to the general (say M-fold) case is quite straightforward. The correct zero-order eigenfunctions, v_{0m} can be written as linear combinations of the original eigenfunctions u_{0k} according to

$$v_{0m} = \sum_{k=1}^{M} C_{mk} u_{0k} \tag{7.38}$$

and the first-order corrections to the energy are obtained from the determinantal equation

$$\begin{vmatrix} H'_{11} - E_1 & H'_{12} & \cdots & H'_{1M} \\ H'_{21} & H'_{22} - E_1 & \cdots & H'_{2M} \\ \vdots & \vdots & & \vdots \\ H'_{M1} & H'_{M2} & \cdots & H'_{MM} - E_1 \end{vmatrix} = 0 \tag{7.39}$$

The coefficients C_{mk} are obtained by substituting the resulting values of E_1 into the generalized form of (7.36) and applying the normalization condition. Once the appropriate set of unperturbed eigenfunctions has been found, higher-order corrections can be applied, just as in the non-degenerate case. However, in many cases, the results of physical interest emerge from the first-order treatment just described.

7.2.1 Nearly degenerate systems

Cases sometimes arise where the unperturbed energies of two or more states are nearly, but not exactly equal. This means that, although the relevant terms in the perturbation expansion are not infinite, they can be very large and the perturbation

series does not converge rapidly, if at all. However, the procedures just described are valid only if the states are actually degenerate.

We will show that a procedure similar to that for the fully degenerate case can be applied in the nearly-degenerate situation. Although perturbation theory cannot be directly applied, we can be guided by the perturbation approach. If we re-examine the procedure described earlier, we see this amounts to using a matrix representation for the part of the system that is described by the states u_1 and u_2. In the degenerate case, we can do this using the perturbation $\hat{H}'$ only because the unperturbed energies of the two states are equal, but in the more general case, we must use the whole Hamiltonian, $\hat{H} = \hat{H}_0 + \hat{H}'$. Instead of (7.36) we get

$$\left.\begin{aligned}(H_{11} - E)C_1 + H_{12}C_2 = 0\\ H_{21}C_1 + (H_{22} - E)C_2 = 0\end{aligned}\right\} \tag{7.40}$$

where

$$H_{11} = \int u_1^*(\hat{H}_0 + \hat{H}')u_1\,d\tau = E_{01} + H'_{11}$$

and

$$H_{12} = \int u_1^*(\hat{H}_0 + \hat{H}')u_2\,d\tau = H'_{12}$$

with similar expressions for H_{21} and H_{22}

The equivalent of (7.36) is then

$$\left.\begin{aligned}(H'_{11} + E_{01} - E)C_1 + H'_{12}C_2 = 0\\ H'_{21}C_1 + (H'_{22} + E_{02} - E)C_2 = 0\end{aligned}\right\} \tag{7.41}$$

The similarity to the fully degenerate case is emphasized, if we express E_{01} and E_{02} in terms of the average energy $\bar{E}_0 = \frac{1}{2}(E_{01} + E_{02})$ and the difference $\Delta E_0 = (E_{01} - E_{02})$; we also define E_1 so that $E = \bar{E}_0 + E_1$. (7.41) then becomes

$$\left.\begin{aligned}(H'_{11} + \tfrac{1}{2}\Delta E_0 - E_1)C_1 + H'_{12}C_2 = 0\\ H'_{21}C_1 + (H'_{22} - \tfrac{1}{2}\Delta E_0 - E_1)C_2 = 0\end{aligned}\right\} \tag{7.42}$$

which is identical to (7.36) in the fully degenerate case, when $\Delta E_0 = 0$. Once the solutions to (7.42) have been obtained, further corrections can be found using the procedure for non-degenerate systems.

This procedure can be extended to systems where more than two states are nearly degenerate. The criterion for near-degeneracy is that the difference between the unperturbed energies is less than the magnitude of the matrix element connecting them. That is

$$\Delta E_0^2 < |H'_{12}|^2 = H'_{12}H'_{21} \tag{7.43}$$

An example of a nearly degenerate system is the one-dimensional metal discussed in example 7.4.

Example 7.3 The Stark effect in hydrogen The effect of applying an electric field to the hydrogen atom in its ground state was discussed earlier in this chapter where we showed that there was no first-order change in the energy of this state. We now consider the first excited state which is fourfold degenerate[1] in the absence of any perturbation and we shall find that this level is split into three when an electric field is applied. This is known as the *Stark effect*.

The four degenerate states corresponding to the unperturbed first excited state of hydrogen all have quantum number n equal to 2 and the eigenfunctions, referred to spherical polar coordinates (r, θ, ϕ), are—cf. chapter 3, equation (3.70) with $z = 1$:

$$\left.\begin{aligned} u_1 &\equiv u_{200} = (8\pi a_0^3)^{-1/2}(1 - r/2a_0)e^{-r/2a_0} \\ u_2 &\equiv u_{210} = (8\pi a_0^3)^{-1/2}(r/2a_0)\cos\theta e^{-r/2a_0} \\ u_3 &\equiv u_{211} = -(\pi a_0^3)^{-1/2}(r/8a_0)\sin\theta e^{i\phi}e^{-r/2a_0} \\ u_4 &\equiv u_{21-1} = (\pi a_0^3)^{-1/2}(r/8a_0)\sin\theta e^{-i\phi}e^{-r/2a_0} \end{aligned}\right\} \tag{7.44}$$

The perturbation due to the applied field (assumed as usual to be in the z direction) is represented by $\hat{H}'$ where—cf. (7.26)

$$\hat{H}' = e\mathcal{E}z = e\mathcal{E}r\cos\theta \tag{7.45}$$

If we now proceed to evaluate the matrix elements we find that most of them vanish because of symmetry. Thus $H'_{11} = H'_{22} = H'_{33} = H'_{34} = H'_{43} = H'_{44} = 0$ because in each case the integrand is antisymmetric in z, and $H'_{13} = H'_{14} = H'_{23} = H'_{24} = H'_{31} = H'_{32} = H'_{41} = H'_{42} = 0$ because these integrands all contain the factor $e^{i\phi}$ and the integration with respect to ϕ is from $\phi = 0$ to $\phi = 2\pi$. This leaves only the matrix elements H'_{12} and H'_{21} which are given by

$$\begin{aligned} H'_{12} &= H'_{21} \\ &= \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty}(8\pi a_0^3)^{-1}(r/2a_0)(1 - r/2a_0) \\ &\quad \times e\mathcal{E}r\cos^2\theta e^{-r/a_0}r^2\,dr\,\sin\theta\,d\theta\,d\phi \\ &= \frac{e\mathcal{E}}{8a_0^4}\int_0^{\pi}\cos^2\theta\sin\theta\,d\theta\int_0^{\infty}(r^4 - r^5/2a_0)e^{-r/a_0}\,dr \\ &= -3e\mathcal{E}a_0 \end{aligned} \tag{7.46}$$

The determinantal equation (7.39) therefore becomes

$$\begin{vmatrix} -E_1 & -3e\mathcal{E}a_0 & 0 & 0 \\ -3e\mathcal{E}a_0 & -E_1 & 0 & 0 \\ 0 & 0 & -E_1 & 0 \\ 0 & 0 & 0 & -E_1 \end{vmatrix} = 0 \tag{7.47}$$

[1] For the purposes of this discussion we shall assume that the perturbation associated with the applied electric field is much larger than the splitting resulting from spin–orbit coupling discussed in the previous chapter so that the latter effect can be ignored.

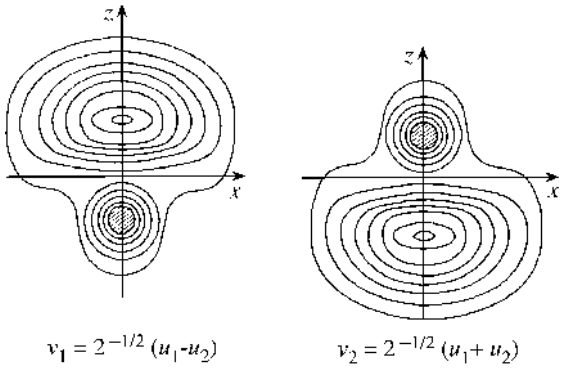


Figure 7.1. A section at $y = 0$ through the position probability distribution corresponding to the $n = 2$ energy state of a hydrogen atom subject to an electric field in the z direction. (A number of contours have been omitted in the high peaks represented by the cross-hatched areas.)

leading to the following expressions for the first-order corrections to the energy and for the zero-order eigenfunctions—making use of the generalized form of (7.36)

$$\left.\begin{array}{ll} E_1 = 3e\mathcal{E}a_0 & v_1 = \dfrac{1}{\sqrt{2}}(u_1 - u_2) \\ E_1 = -3e\mathcal{E}a_0 & v_2 = \dfrac{1}{\sqrt{2}}(u_1 + u_2) \\ E_1 = 0 & v_3 = u_3 \qquad \text{and} \qquad v_4 = u_4 \end{array}\right\} \tag{7.48}$$

We should note several points about these results. First, the degeneracy of the last two states has not been lifted by the perturbation so any linear combination of u_3 and u_4 is a valid eigenfunction with $E_1 = 0$. Secondly, we see from figure 7.1, which shows a plot of the probability densities $|v_1|^2$ and $|v_2|^2$, that these are not symmetric across the plane $z = 0$, implying that in these states the atom possesses a dipole moment that is aligned antiparallel to the field in the first case and parallel to it in the second. The energy changes can therefore be thought of as arising from the interaction between these dipoles and the applied field. However, in contrast to the polarization of the ground state discussed earlier (section 7.1) these dipole moments are not generated by the physical operation of the field and their creation does not require the expenditure of any energy: the polarized states are simply two of the possible eigenfunctions of the degenerate unperturbed system. The energy change is therefore proportional to the field magnitude $\mathcal{E}$, rather than to $\mathcal{E}^2$. Thirdly, we see from (7.48) that neither v_3 nor v_4 (and also by implication no linear combination of these functions) possesses a dipole moment and therefore neither has a first-order interaction with the field; the energy of these states is not affected by the perturbation to first order and they remain degenerate even in the

presence of the field. Finally, we note that the occurrence of the Stark effect is a result of a mixing of the wavefunctions associated with the spherically symmetric $2s$ state and the $2p$ state with $m = 0$. In the hydrogenic atom these states are 'accidentally' degenerate but this is not the case in other atomic systems where a linear Stark effect is therefore not found. Apart from hydrogen, a linear splitting is observed only in the case of a few atoms such as helium where the energy difference between the s and p states is small compared with the perturbation resulting from a strong electric field (cf. chapter 10).

Example 7.4 Electrons in a one-dimensional solid The fact that metals can carry electric currents implies that they contain electrons that are mobile and not attached to particular atoms; such electrons are known as 'free electrons'. In many common metals (e.g. sodium) all but one of the atomic electrons are tightly bound in 'closed shells', while the remaining electron is effectively free (see chapter 10). In this example we shall develop a one-dimensional model of a solid in which we study the behaviour of otherwise free electrons in a potential similar to that arising from the nuclei and other electrons in a real solid. We shall assume that this potential is weak enough to be treated as a perturbation and we shall see that this simple model provides quite a good explanation of the electrical conductivity of solids.

The atoms in a crystalline solid are arranged on a regular lattice, so, in one dimension, we expect the potential experienced by a free electron to vary periodically with distance. If we call this repeat distance a, the simplest form of such a periodic potential is

$$V(x) = V_0 \cos(2\pi x/a) \tag{7.49}$$

V_0 is assumed to be small so that V can be treated as a perturbation, where the unperturbed eigenfunctions correspond to completely free electrons and therefore have the form

$$u_{0k} = L^{-1/2} \exp(ikx) \tag{7.50}$$

where the factor $L^{-1/2}$ ensures that the eigenfunctions are normalized when integrated over the distance L, which is taken as the length of the macroscopic piece of solid being considered. If this contains N atoms, it follows that $L = Na$. The unperturbed energies E_{0k} are given by (cf. section 2.4)

$$E_{0k} = \frac{\hbar^2 k^2}{2m} \tag{7.51}$$

The matrix elements representing the linking of the states u_{0k} and $u_{0k'}$ are given

by

$$\begin{aligned} H'_{kk'} &= L^{-1}\int_0^{Na} \exp(-ikx)V_0\cos(2\pi x/a)\exp(ik'x)\,dx \\ &= \frac{V_0}{2L}\int_0^{Na}[\exp i(k'-k+2\pi/a)x + \exp i(k'-k-2\pi/a)x]\,dx \\ &= \frac{V_0}{2} \qquad \text{if } k'-k = \pm\frac{2\pi}{a} \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{7.52}$$

We first note that the diagonal elements H'_{kk} are all zero. If the states u_{0k} and $u_{0k'}$ are degenerate, $k' = -k$ and the matrix element connecting these states is non-zero only if $k = \pm\pi/a$. The first-order changes to the energies of these state are then given by equation (7.39) which now has the form

$$\begin{vmatrix} -E_1 & \frac{1}{2}V_0 \\ \frac{1}{2}V_0 & -E_1 \end{vmatrix} = 0 \tag{7.53}$$

Thus $E_1 = \pm\frac{1}{2}V_0$ and the degeneracy of those states where $k = \pm\pi/a$ is lifted in first order.

We now consider states with other values of k. If k is very different from $\pm\pi/a$, so that the unperturbed energy difference $(E_{0k} - E_{0k'})$ (where $k - k' = \pm 2\pi/a$) which appears in the denominator of (7.21) is large compared with $\frac{1}{2}V_0$, the second-order term will be small. However, if the values of k are in the near vicinity of, say π/a, the matrix element of the perturbation linking this state with that with $k' \sim -\pi/a$ will be much greater than the difference between the unperturbed energies of the states k and k', and we can treat the problem as an example of a 'nearly degenerate' system discussed earlier.[2] Referring to (7.42), the determinantal equation is now

$$\begin{vmatrix} \frac{1}{2}\Delta E_0 - E_1 & \frac{1}{2}V_0 \\ \frac{1}{2}V_0 & -\frac{1}{2}\Delta E_0 - E_1 \end{vmatrix} = 0 \tag{7.54}$$

which leads directly to

$$E_1 = \pm\tfrac{1}{2}(\Delta E_0^2 + V_0^2)^{1/2} \tag{7.55}$$

which agrees with the value of $\pm\frac{1}{2}V_0$ obtained earlier for the degenerate case where $k = \pi/a$. Remembering the definitions of ΔE_0 and E_1 and using (7.51) we have

$$E = \frac{\hbar^2}{4m_e}\left[k^2 + \left(k - \frac{2\pi}{a}\right)^2\right] \pm \frac{1}{2}\left(\left[\frac{\hbar^2}{4m_e}\left[k^2 - \left(k - \frac{2\pi}{a}\right)^2\right]\right]^2 + V_0^2\right)^{1/2} \tag{7.56}$$

[2] The only other state linked by the perturbation to that with $k \sim \pi/a$ is one with $k \sim 3\pi/a$ and the energy difference between these two is large.

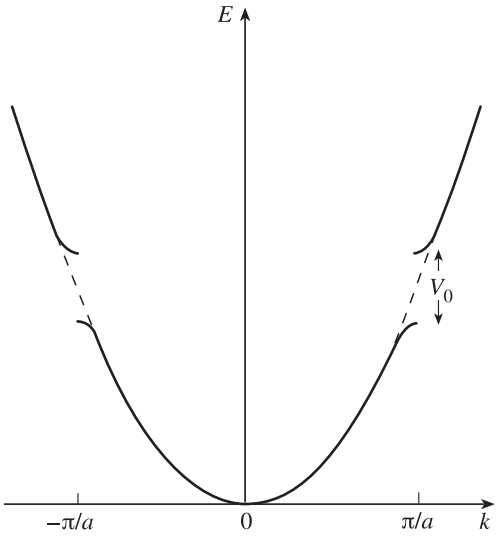


Figure 7.2. The energy of an electron in a one-dimensional metal as a function of wave number k, compared with that of a free electron (broken line), showing the energy gaps at $k = \pm\pi/a$.

As k moves away from π/a, the energy should approach its unperturbed value. This is achieved if the positive sign in (7.56) is assigned to the state with wavevector k if $k > \pi/a$ and the negative sign when $k < \pi/a$. This implies that the negative sign corresponds to the state $k - 2\pi/a$ when this quantity is less than $-\pi/a$ and vice versa. In this way, we obtain the curve of E_k as a function of k shown in figure 7.2. A notable feature of this diagram is the appearance of gaps of width V_0 in the energy spectrum corresponding to the points $k = \pm\pi/a$. The influence of these gaps on the physical properties of the system is considerable and we shall discuss this shortly. First, however, we must consider more carefully the boundary conditions to be imposed on the problem.

The length L was previously described as the length of the piece of solid under consideration so the most obvious boundary condition would be to require the wavefunction to be zero outside the range between $x = 0$ and $x = L$. Such 'fixed boundary conditions' can be used, but they give rise to considerable mathematical complications associated with the fact that plane waves of the form (7.50) are no longer eigenfunctions of the unperturbed system. It is therefore more convenient to use 'periodic boundary conditions' where we impose the condition that $u_k(x) = u_k(x + L)$ and the allowed values of k are therefore those where $k = 2n\pi/L$, n being an integer. This boundary condition corresponds physically to considering the one-dimensional line to be bent round into a closed loop, when the allowed values of k follow from the condition that the wavefunction

be single-valued. As bulk properties such as the electrical conductivity of real solids are not dependent on parameters such as the size and shape of the sample being considered, it is a reasonable assumption that the precise form of boundary conditions will not be important when discussing similar properties of our one-dimensional model. The most convenient form can therefore be chosen and this turns out to correspond to periodic boundary conditions.

We previously saw that the number of atoms associated with a length L of this one-dimensional solid is N where $L = Na$ and a is the repeat distance of the periodic potential. The number of energy states associated with values of k between plus and minus π/a is also N, because the periodic boundary conditions require that successive allowed values of k are separated by $2\pi/L$. It follows from the Pauli exclusion principle to be discussed in chapter 10 that each state can be occupied by no more than two electrons, which must have opposite spin. In the case where each atom contributes one free electron to the system, the ground state of the system will therefore correspond to the $(N/2)$ states of lowest energy being filled. If now an electric field is applied, some of the electrons will be excited so that there are more with (say) positive k than with negative k and, as k is proportional to the electron momentum, an electric current results. In this case the one-dimensional solid behaves like a metal. However, if there are two free electrons per atom, all the states in the band of energies whose k values lie between plus and minus π/a will be occupied. If V_0 is large enough, an applied field will be unable to excite any electrons, and there will therefore be as many electrons with positive k as there are with negative k; no current can then flow, and the one-dimensional solid in this case is an insulator. If, however, V_0 (and consequently the size of the energy gap) is small enough, some electrons will be thermally excited into states with $|k|$ greater than π/a where they will be mobile and can respond to a field. This excitation leaves behind 'holes' in the otherwise full band and it can be shown that these have the same electrical properties as positively charged mobile particles. Such a system is a one-dimensional intrinsic semi-conductor.

This discussion can be generalized to the case where there are more than two electrons per atom if we remember that a general periodic potential has Fourier components with repeat distances a, $a/2$ etc., giving rise to energy gaps at $k = \pm\pi/a, \pm 2\pi/a$, etc. It is then clear that any one-dimensional solid which possesses an odd number of free electrons per atom will be a metal, while an even number will imply insulating or semi-conducting properties. The arguments can be extended to three-dimensional solids where the rule is not so simple, but where quantum mechanics has been successfully used to explain the wide range of physical phenomena displayed by such materials. The interested reader should consult a textbook on solid state physics (for example, J. R. Hook and H. E. Hall *Solid State Physics*, Wiley, New York, 1991).

7.3 The variational principle

This approximate method can be applied to eigenvalue problems where we know the operator $\hat{H}$ and can make some guess as to the form of the eigenfunction corresponding to the lowest energy state of the system. The variational principle can then be used to improve our guessed eigenfunction and produce a maximum value, or upper bound, to the ground-state energy. The usefulness of the variational principle depends strongly on our ability to make a good initial guess, and the symmetry and other physical properties of the system can often be useful guides to this. The variational principle can be extended to consider states other than the ground state of the system, but this has limited applications and will not be discussed here. We shall now proceed to describe the method in detail.

Let v be a function that is to approximate to the ground-state eigenfunction of the Hamiltonian operator $\hat{H}$. We shall now show that the expectation value of $\hat{H}$ calculated using v cannot be less than the true ground-state eigenvalue E_0. Allowing for the fact that v may not be normalized, this expectation value $\langle \hat{H} \rangle$ is given by

$$\langle \hat{H} \rangle = \frac{\int v^* \hat{H} v \, d\tau}{\int v^* v \, d\tau} \tag{7.57}$$

Using completeness, v can be expressed as a linear combination of the true, but unknown, eigenfunctions of $\hat{H}$:

$$v = \sum_n a_n u_n \tag{7.58}$$

Using (7.58) the numerator of (7.57) can be expressed as

$$\begin{aligned} \int v^* \hat{H} v \, d\tau &= \int \left(\sum_n a_n^* u_n^* \right) \hat{H} \left(\sum_m a_m u_m \right) d\tau \\ &= \sum_{nm} a_n^* a_m \int u_n^* \hat{H} u_m \, d\tau \\ &= \sum_{nm} a_n^* a_m E_m \delta_{nm} \\ &= \sum_n |a_n|^2 E_n \end{aligned} \tag{7.59}$$

using orthonormality. If E_0 is the ground-state eigenvalue it must be smaller than all the other E_n's, so we can write

$$\int v^* \hat{H} v \, d\tau \geqslant E_0 \sum_n |a_n|^2 \tag{7.60}$$

The denominator of (7.57) can be similarly shown to be given by

$$\int v^* v \, d\tau = \sum_n |a_n|^2 \tag{7.61}$$

and combining (7.57), (7.60), and (7.61) we get

$$\langle \hat{H} \rangle \geqslant E_0 \tag{7.62}$$

as required. Thus by guessing an approximate eigenfunction, v, we can never get an expectation value of the energy that is less than the true value. Hence we can calculate an upper bound to the ground-state energy. Moreover, if v contains adjustable parameters, these can be varied until $\langle \hat{H} \rangle$ has its minimum value, when v will represent the best possible approximation of this form.

Example 7.5 The harmonic oscillator In this example we suppose that we do not know the ground-state eigenfunction for a one-dimensional harmonic oscillator, but have guessed that it is similar to that for a particle in an infinite potential well. That is,

$$\left.\begin{aligned} v &= a^{-1/2}\cos(\pi x/2a) \qquad a \leqslant x \leqslant a \\ &= 0 \qquad\qquad\qquad\quad |x| > a \end{aligned}\right\} \tag{7.63}$$

We wish to find the value of a for which the expectation value of the energy is a minimum. Using (7.57)

$$\langle \hat{H} \rangle = a^{-1} \int_{-a}^{a} \cos(\pi x/2a) \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2} m\omega^2 x^2 \right) \cos(\pi x/2a)\, dx$$

where m is the particle mass and ω the classical frequency of the oscillator. The differentiation and integration are straightforward, if a little tedious, and we get

$$\langle \hat{H} \rangle = \frac{\hbar^2 \pi^2}{8ma^2} + m\omega^2 a^2 \left(\frac{1}{6} - \frac{1}{\pi^2} \right) \tag{7.64}$$

To find an expression for a corresponding to the minimum value of $\langle \hat{H} \rangle$ we put $\partial \langle \hat{H} \rangle / \partial a = 0$ and get

$$-\frac{\hbar^2 \pi^2}{4ma^3} + 2m\omega^2 a \left(\frac{1}{6} - \frac{1}{\pi^2} \right) = 0$$

so that

$$\begin{aligned} a &= \pi \left[\frac{3}{4(\pi^2 - 6)} \right]^{1/4} \left(\frac{\hbar}{m\omega} \right)^{1/2} \\ &= 2.08 \left(\frac{\hbar}{m\omega} \right)^{1/2} \end{aligned} \tag{7.65}$$

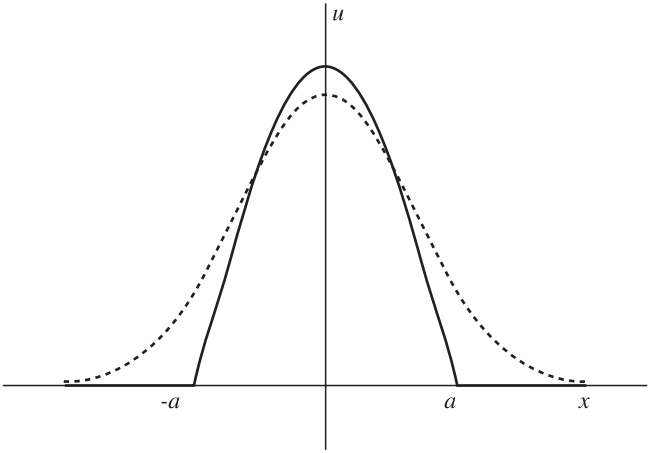


Figure 7.3. The wavefunction corresponding to the ground state of a one-dimensional harmonic oscillator (broken line) compared with that for an infinite-sided well whose parameters are chosen using the variational principle (continuous line).

We can substitute from (7.65) into (7.64) to obtain the minimum value of $\langle \hat{H} \rangle$ which we write as $\langle \hat{H} \rangle_{\min}$. Thus

$$\begin{aligned} \langle \hat{H} \rangle_{\min} &= \frac{1}{2} \left(\frac{\pi^2 - 6}{3} \right)^{1/2} \hbar\omega \\ &= 0.568 \hbar\omega \end{aligned} \tag{7.66}$$

Thus we have been able to set an upper bound to the energy of the oscillator which is within 14% of the true value of $\frac{1}{2}\hbar\omega$. Figure 7.3 compares the approximate wavefunction evaluated using the value of a given in (7.65) with the exact ground-state eigenfunction obtained in chapter 2.

Example 7.6 The atomic polarizability of hydrogen This problem was previously treated by perturbation theory where we showed (7.33) that an upper bound to the polarizability α is $64\pi a_0^3/3$. As the energy of an atom in an electric field of magnitude $\mathcal{E}$ is equal to $-\frac{1}{2}\varepsilon_0 \alpha \mathcal{E}^2$, this corresponds to a lower bound for the energy; we shall now use the variational principle to set an upper bound for the energy and hence a lower bound to α.

We showed earlier—cf. (7.26)—that the Hamiltonian for a hydrogen atom in a uniform electric field $\mathcal{E}$ in the positive z direction is

$$\hat{H} = \hat{H}_0 + e\mathcal{E}z \tag{7.67}$$

We expect the application of an electric field to polarize the atom, which means that the negatively-charged electron will be pulled in a direction opposite to the field, while the positively charged nucleus is pushed parallel to it. The wavefunction should therefore be such as to enhance the probability of finding the electron at negative rather than positive values of z. A simple (unnormalized) function with these properties is

$$v = u_0(1 - \beta z) \tag{7.68}$$

where u_0 is the ground-state eigenfunction of $\hat{H}_0$ given in chapter 3 (3.70) and β is a constant. We note that u_0 is a real function, and we also assume β to be real. We can therefore write the expectation value of $\hat{H}$ as

$$\begin{aligned}\langle \hat{H} \rangle &= \frac{\int u_0(1-\beta z)(\hat{H}_0 + e\mathcal{E}z)u_0(1-\beta z)\,d\tau}{\int u_0^2(1-\beta z)^2\,d\tau} \\ &= \left[\int u_0 \hat{H}_0 u_0\,d\tau - \beta\int u_0(z\hat{H}_0 + \hat{H}_0 z)u_0\,d\tau + e\mathcal{E}\int u_0 z u_0\,d\tau \right. \\ &\quad \left. + \beta^2\int u_0 z\hat{H}_0 z u_0\,d\tau - 2\beta e\mathcal{E}\int u_0 z^2 u_0\,d\tau + \beta^2 e\mathcal{E}\int u_0 z^3 u_0\,d\tau\right] \\ &\quad \div \left[\int u_0^2\,d\tau - 2\beta\int u_0 z u_0\,d\tau + \beta^2\int u_0 z^2 u_0\,d\tau\right]\end{aligned} \tag{7.69}$$

Remembering that u_0 is the ground-state energy eigenfunction in the zero-field case, which is spherically symmetric, all integrals that involve odd powers of z, are equal to zero on symmetry grounds. Moreover, $\int u_0 z\hat{H}_0 z u_0\,d\tau$ also vanishes, as can be shown by substituting the spherical polar expressions for the operators and eigenfunctions (cf. chapter 3), while the integral $\int u_0 z^2 u_0\,d\tau$ was earlier shown (7.31) to be equal to a_0^2. Equation (7.69) therefore becomes

$$\begin{aligned}\langle \hat{H} \rangle &= \frac{E_0 - 2e\mathcal{E}\beta a_0^2}{1 + \beta^2 a_0^2} \\ &\simeq E_0(1 - \beta^2 a_0^2) - 2e\mathcal{E}\beta a_0^2\end{aligned} \tag{7.70}$$

where we have ignored powers of β higher than the second as we are interested in the case of low fields where β is expected to be small. We can now differentiate to find a value of β corresponding to the minimum value of $\langle \hat{H} \rangle$:

$$-2a_0^2\beta E_0 - 2e\mathcal{E}a_0^2 = 0$$

that is,

$$\beta = -\frac{e\mathcal{E}}{E_0} \tag{7.71}$$

Substituting into (7.70), the minimum value $\langle \hat{H} \rangle_{\min}$ of $\langle \hat{H} \rangle$ is

$$\begin{aligned} \langle \hat{H} \rangle_{\min} &= E_0 + \frac{e^2 \mathcal{E}^2}{E_0} a_0^2 \\ &= E_0 - 8\pi \varepsilon_0 a_0^3 \mathcal{E}^2 \end{aligned} \tag{7.72}$$

using (3.72). This represents an upper bound to the ground-state energy, so the corresponding lower bound for the polarizability is $16\pi a_0^3$. We can combine this with the upper bound obtained by perturbation methods and the Unsöld approximation (7.33) to give

$$16\pi a_0^3 \leqslant \alpha \leqslant 64\pi a_0^3/3 \tag{7.73}$$

Taking as a best estimate, the average of these two bounds, we get

$$\alpha = (18.8 \pm 2.8)\pi a_0^3 \tag{7.74}$$

and we have therefore obtained a theoretical value for the polarizability that is rigorously correct to within about 14%. Moreover, it is in good agreement with experimental value of $18.4\pi a_0^3$.

A combination of perturbation and variational calculations can quite often be used in a similar way to that just described to obtain upper and lower bounds to the theoretical estimates of physical quantities that cannot be calculated directly—perhaps because the unperturbed eigenvalues and eigenfunctions are unknown. Such methods are frequently used, for example, in the calculation of the polarizabilities of non-hydrogen atoms and of molecules, and in estimating the 'van der Waals' interactions between atoms in gases. In favourable cases, by the use of sophisticated trial functions in the variational method and special techniques to estimate the sums over states in the perturbation expressions, the upper and lower bounds can be made to approach each other very closely so that highly accurate theoretical results can be obtained.

Problems

7.1 A particle moves in a potential given by

$$V = V_0 \cos(\pi x/2a) \quad (-a \leqslant x \leqslant a) \qquad V = \infty \quad (|x| > a)$$

where V_0 is small. Treat this problem as a perturbation on the case of a particle in an infinite-sided square well of length $2a$ and calculate the changes in the energies of the three lowest energy states to first order in V_0.

7.2 A particle moves in a potential given by

$$V = V_0 \quad (-b \leqslant x \leqslant b) \qquad V = 0 \quad (b < |x| \leqslant a) \qquad V = \infty \quad |x| > a$$

Calculate the energies of the three lowest states to first order in V_0 using a similar procedure to that in problem 7.1.

7.3 The hydrogenic atom was treated in chapter 3 on the assumption that the nucleus has a point charge, while in reality the nuclear charge is spread over a small volume. Show that if we were

to assume that the nuclear charge was in the form of a thin shell of radius δ, where $\delta \ll a_0$, the ground -state energy of a hydrogen atom would be increased by an amount equal to $Ze^2\delta^2/(6\pi\varepsilon_0 a_0^3)$. Calculate this quantity for hydrogen assuming $\delta = 10^{-15}$ m and express your result as a fraction of the total ground-state energy.

7.4 A particle of mass m is attached by a massless rigid rod of length a to a fixed point and can rotate about a fixed axis passing through this point. Its energy eigenvalues and corresponding eigenfunctions are then $E_n = \hbar^2 n^2/2ma^2$ and $u_n = (2\pi)^{-1/2}\exp(in\phi)$ respectively (cf. problem 5.2). If this system is now perturbed by a potential $V_0\cos(2\phi)$, calculate the first-order changes in the three lowest energy levels, remembering to allow for degeneracy where appropriate. Show that the second-order change in the ground ($n = 0$)-state energy is equal to $-ma^2V_0^2/4\hbar^2$.

7.5 Show that a first-order Stark effect is possible only if the unperturbed eigenfunctions do not have a definite parity. How then is it possible to observe the Stark effect in hydrogen where the potential is centrosymmetric?

7.6 Use perturbation theory to show that the first-order changes in energy levels due to spin–orbit coupling are given by

$$[j(j+1) - l(l+1) - s(s+1)]\hbar^2\langle f(r)\rangle$$

in the notation of chapter 6 where the expectation value $\langle f(r)\rangle$ is calculated using only the radial part of the unperturbed eigenfunction. Use the fact that the splitting of the D lines in sodium is about 6×10^{-10} m and that their mean wavelength is 5.9×10^{-7} m to estimate $\langle f(r)\rangle$ in the case of the $l = 1$ state involved in the D lines of sodium.

7.7 Use the result obtained in problem 7.6 to show that spin–orbit coupling splits the $2p$ level in hydrogen into a doublet whose separation, expressed as a fraction of the mean unperturbed energy, is equal to $(e^2/8\pi\varepsilon_0\hbar c)^2$.

7.8 Use perturbation theory to obtain the spin–orbit correction to the strong-field Zeeman levels (6.63). *Hint*. First show that the perturbation can be written in the form $2f(r)\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}$; remember that the unperturbed eigenfunctions are eigenfunctions of $\hat{L}_z$ and $\hat{S}_z$.

7.9 Use the variational principle and the trial function $\exp(-\alpha x^2)$ to obtain an upper limit for the ground-state energy of a one-dimensional harmonic oscillator. Compare your results with the exact expressions given in chapter 2.

Chapter 8

Time dependence

So far in our development of quantum mechanics, we have given very little consideration to problems involving the time dependence of the wavefunction. This is rather surprising as in classical mechanics it is time-dependent phenomena—i.e. dynamics rather than statics—that command most attention. Moreover, most experimental observations necessarily involve some change in the quantity being observed, so we might expect that the successful prediction of experimental results will require a detailed understanding of the way in which a system changed in time. In fact we have made some implicit assumptions about time dependence and the principal reason why we have got so far without discussing it in detail is that many observed quantum phenomena are associated with sudden discontinuous changes between otherwise stable states. Thus most of our information concerning the energy levels of atoms has been obtained from measurements of the frequencies of electromagnetic radiation emitted or absorbed as the atom undergoes a transition from one energy eigenstate to another, assuming the correctness of the formula $E = \hbar\omega$, but not considering the mechanism of the transition in any detail. Assumptions concerning time dependence are also implicit in the quantum theory of measurement which refers to the probability of obtaining a particular result *following a measurement* performed on a system in a given state; thus, for example, the state vector of a spin-half particle changes from being an eigenvector $\hat{S}_z$ to being one of $\hat{S}_x$ following a measurement of the latter property using an appropriately oriented Stern–Gerlach apparatus. As was pointed out in chapter 4, and will be discussed further in chapters 12 and 13, the 'collapse' of the wavefunction associated with a measurement is not a consequence of the time-dependent Schrödinger equation.

There are, however, a number of problems in which the time dependence of the wavefunction must be considered explicitly, and some of these will be discussed in the present chapter. We already know from postulate 4.5 that the basic equation governing the time evolution of the wavefunction between

measurements is the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi \tag{8.1}$$

where $\hat{H}$ is the Hamiltonian operator representing the total energy of the system. We shall initially consider the case where $\hat{H}$ is not itself explicitly time dependent. We later turn to the more general case and consider the 'sudden approximation', where $\hat{H}$ changes suddenly from one time-independent form to another; and 'time-dependent perturbation theory', where the time dependence is confined to a part of the Hamiltonian which can be considered to be small. This will enable us to solve the problem of an atom subject to a time-varying field and will lead to an understanding of the occurrence of transitions between different energy states. We shall also see why some of these transitions are more probable, and therefore associated with more intense spectral lines, than others. We shall then discuss the Ehrenfest theorem which is used to clarify the connection between quantum and classical mechanics. We close the chapter with a discussion of the application of the general results to the particular example of the ammonia maser.

8.1 Time-independent Hamiltonians

This case was considered briefly in the discussion following postulate 4.5 where it was shown that, if the wavefunction at some initial time $t = 0$ is given by $\Psi(\mathbf{r}, 0)$ where

$$\Psi(\mathbf{r}, 0) = \sum_n a_n(0)u_n(\mathbf{r}) \tag{8.2}$$

and the $u_n(\mathbf{r})$ are the energy eigenfunctions, i.e.

$$\hat{H}u_n = E_n u_n \tag{8.3}$$

then the wavefunction $\Psi(\mathbf{r}, t)$ at time t is given by

$$\Psi(\mathbf{r}, t) = \sum_n a_n(0)u_n(\mathbf{r})\exp(-iE_n t/\hbar) \tag{8.4}$$

We pointed out that (8.4) contains the quantum-mechanical equivalent of the conservation of energy: if the energy has once been measured so that the wavefunction is an energy eigenfunction, it retains this form indefinitely (apart from a time-dependent phase factor) and any subsequent energy measurement is then certain to produce the same result. Equation (8.4) can also be used to study the behaviour of systems that are not initially in energy eigenstates, and we shall now consider two particular examples that illustrate this point.

The harmonic oscillator

The energy levels of a one-dimensional harmonic oscillator were shown in chapter 2 to be $E_n = (n + \frac{1}{2})\hbar\omega$ where ω is the classical angular frequency. We can use this expression and the general theory described earlier to study the behaviour of such an oscillator when it is not initially in an energy eigenstate. Equation (8.4) becomes, in this case,

$$\Psi(x, t) = \sum_n a_n(0)u_n(x) \exp\left(-i(n + \tfrac{1}{2})\omega t\right) \tag{8.5}$$

If we consider the particular time T corresponding to the classical period of the oscillator (that is, $T = 2\pi/\omega$) we get

$$\begin{aligned}\Psi(x, T) &= \sum_n a_n(0)u_n(x) \exp\left(-i(n + \tfrac{1}{2})2\pi\right) \\ &= -\sum_n a_n(0)u_n(x) \\ &= -\Psi(x, 0)\end{aligned} \tag{8.6}$$

Thus, whatever the initial conditions, the wavefunction at time T is equal to minus that at time zero. It clearly follows that at time $2T$ the wavefunction will be identical to that at time zero, so the wavefunction varies periodically with a frequency half that of the classical oscillator. However, it should be noted that the sign of the wavefunction has no physical significance, and we conclude that all the physical properties of the harmonic oscillator will vary periodically in time with a frequency identical to the classical frequency. Figure 8.1 illustrates this by showing the time evolution of the position probability density associated with a wavefunction whose form at $t = 0$ is the normalized sum of the two lowest energy eigenfunctions.

We can use this example as a further illustration of the correspondence principle, previously discussed in connection with the harmonic oscillator in chapter 2. This states that the results of quantum mechanics go over to those of classical mechanics in the limit where the total energy is large compared with the separation of the states. A classical oscillator, such as a clock pendulum, appears always to have a well-defined position which varies sinusoidally with the classical frequency; quantum mechanically this means that the initial wavefunction must have the form of a narrow localized pulse. It follows from this that this pulse will move from side to side with the classical frequency and amplitude. This implies that a large number of energy eigenfunctions must contribute to the expansion (8.5). There must therefore be some uncertainty in the value of the energy, and the classical limit is reached when the average energy is so large that this uncertainty is an undetectably small fraction of the total.

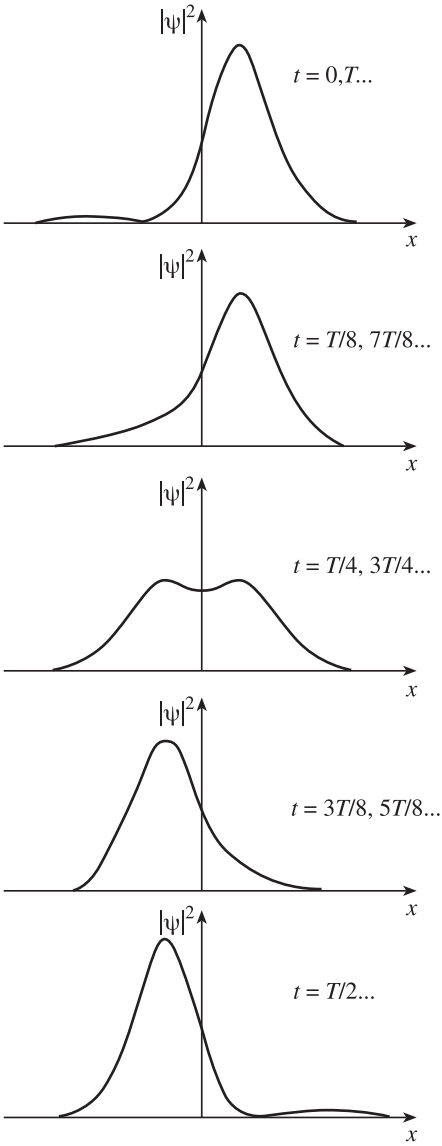


Figure 8.1. The time evolution of the position probability density corresponding to a particle in a harmonic oscillator potential whose wavefunction at $t = 0$ is the normalized sum of the two lowest energy eigenfunctions.

Spin-half particle in a magnetic field

Although the time-dependent Schrödinger equation (8.1) has been written in terms of the wave representation, it is equally valid if Ψ is replaced by a column vector whose elements are, in general, time dependent, and the Hamiltonian is expressed as a matrix (cf. chapter 6). If we consider the particular case of a spin-half particle with no orbital angular momentum in a magnetic field B which is in the z direction, (8.1) becomes

$$i\hbar\frac{\partial}{\partial t}\begin{bmatrix}a_1(t)\\a_2(t)\end{bmatrix} = \hat{H}\begin{bmatrix}a_1(t)\\a_2(t)\end{bmatrix} \tag{8.7}$$

and the Hamiltonian $\hat{H}$ is (cf. (6.34))

$$\begin{aligned}\hat{H} &= -\boldsymbol{\mu}_{\mathbf{S}}\cdot\mathbf{B}\\ &= \frac{e}{m_e}B\hat{S}_z\\ &= \frac{e\hbar}{2m_e}B\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}\end{aligned} \tag{8.8}$$

where $\hat{S}_z$ has been expressed in terms of a spin-matrix—cf. (6.15) and (6.16). Substituting from (8.8) into (8.7) and expanding we get

$$\left.\begin{aligned}i\hbar\frac{\partial a_1}{\partial t} &= (e\hbar B/2m_e)a_1\\ i\hbar\frac{\partial a_2}{\partial t} &= -(e\hbar B/2m_e)a_2\end{aligned}\right\} \tag{8.9}$$

and therefore

$$\left.\begin{aligned}a_1(t) &= a_1(0)\exp(-\tfrac{1}{2}i\omega_p t)\\ a_2(t) &= a_2(0)\exp(\tfrac{1}{2}\omega_p t)\end{aligned}\right\} \tag{8.10}$$

where $\omega_p = eB/m_e$. If the initial state is an eigenstate of the energy, corresponding to $\hat{S}_z = \frac{1}{2}\hbar$, at $t = 0$, then $a_1(0) = 1$ and $a_2(0) = 0$, and the only change in time is that contained in a phase factor multiplying the whole state vector which has no physical consequences. However, a more interesting case is where the initial state is an eigenstate of $\hat{S}_x$: for example, that where $a_1(0) = a_2(0) = 2^{-1/2}$, corresponding to an eigenvalue of $\frac{1}{2}\hbar$. It follows directly from (8.10) that the wavevector at time t now has the form

$$\frac{1}{\sqrt{2}}\begin{bmatrix}\exp(-\frac{1}{2}i\omega_p t)\\ \exp(\frac{1}{2}\omega_p t)\end{bmatrix} \tag{8.11}$$

This expression is again an eigenvector of $\hat{S}_x$ with the same eigenvalue when $t = 2\pi/\omega_p$, $4\pi/\omega_p$, $6\pi/\omega_p$, etc. Moreover, we can show that at other times

(8.11) is an eigenvector of the operator $\hat{S}_\phi$ representing a measurement of the component of spin in a direction in the xy plane at an angle ϕ to the x axis where $\phi = \omega_p t$. Consider the matrix representing the operator $\hat{S}_\phi$:

$$\begin{aligned}\hat{S}_\phi &= \hat{S}_x \cos\phi + \hat{S}_y \sin\phi \\ &= \frac{1}{2}\hbar \begin{bmatrix} 0 & \exp(-i\phi) \\ \exp(i\phi) & 0 \end{bmatrix}\end{aligned} \tag{8.12}$$

The eigenvector of this matrix with eigenvalue $\frac{1}{2}\hbar$ is

$$\frac{1}{\sqrt{2}} \begin{bmatrix} \exp(-i\phi/2) \\ \exp(i\phi/2) \end{bmatrix} \tag{8.13}$$

and the required result follows directly from a comparison of (8.11) and (8.13). We conclude, therefore, that if the system is initially in an eigenstate of $\hat{S}_x$ with eigenvalue $\frac{1}{2}\hbar$, it will always be in a similar eigenstate of the operator $\hat{S}_\phi$ whose direction rotates in the xy plane with an angular velocity ω_p. It is tempting to conclude from this that the angular-momentum vector of the particle precesses about the field direction with this angular velocity, which is what would happen in the similar classical situation. But it is important not to pursue this analogy too far. In classical precession, the direction of the angular-momentum vector, and hence the magnitudes of all three of its components, always have known values; but in quantum mechanics only one angular-momentum component can be measured at any given time. For example, the precession model would imply that the y component of the angular momentum would be zero at the times when ϕ is zero, but we know from quantum mechanics and experiment that a measurement of this quantity always yields a result equal to either plus or minus $\frac{1}{2}\hbar$ and never zero. The application of this precession model (sometimes known as the 'vector model') can be applied rigorously only in the semi-classical limit where the quantum numbers are large.

Finally we note a further interesting consequence of equation (8.13): if we add 2π to the angle ϕ, that is if we rotate the system through 360°, the sign of the eigenfunction is reversed, and we see from (8.11) that this sign change also occurs under the influence of a magnetic field after a time $2\pi/\omega_p$. Usually this has no effect on the physical properties of the system, for the same reasons as those discussed earlier in connection with the sign of the wavefunction of the harmonic oscillator. However, experiments can be conducted in which the occurrence of the sign change is confirmed. A beam of spin-half particles is directed into an apparatus through which there are two possible paths (see figure 8.2). Each path contains a region in which there is a magnetic field and the two magnetic fields have the same magnitude, but are in opposite directions. The speed of the particles is adjusted so that the magnetic fields rotate the spin by 180° in opposite directions depending on the path followed, and the wavefunctions associated with each path therefore have opposite sign. An interference experiment is performed in which

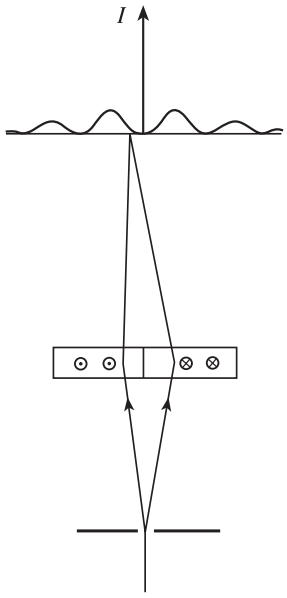


Figure 8.2. Neutrons pass through the lower slit and then enter a region where there is a magnetic field directed either vertically upwards (on the left) or downwards (on the right). The resulting interference pattern has a minimum in the centre if the relative rotation of the spins passing along the two paths is an odd number multiplied by 2π.

we do not know which route the particles have followed so that the wavefunction on the far side of the apparatus is the sum of those associated with the two paths. The resulting diffraction pattern therefore has zero intensity at its centre due to the destructive interference between the oppositely rotated components. In contrast, if $B = 0$, the centre corresponds to a maximum in the interference pattern. Such experiments were carried out by a number of workers in the mid 1970s using neutrons, and their results confirm these theoretical predictions.

8.2 The sudden approximation

So far we have restricted our discussion to systems whose Hamiltonians have no explicit time dependence, but we shall now extend our treatment to include cases where time-varying forces are acting. These problems can often be very difficult to solve and we shall restrict our consideration to those where particular simplifying assumptions can be applied. One such case is the *sudden approximation* which can be used when the Hamiltonian changes instantaneously from one time-independent form—say $\hat{H}_1$—to another—say $\hat{H}_2$—at a time which

we take to be $t = 0$. Thus

$$\left.\begin{aligned} \hat{H} &= \hat{H}_1 \qquad t \leqslant 0 \\ \hat{H} &= \hat{H}_2 \qquad t > 0 \end{aligned}\right\} \tag{8.14}$$

We assume that the eigenfunctions of $\hat{H}_1$ and $\hat{H}_2$ are u_n and v_n respectively and that the system is known to be in one of the eigenstates of $\hat{H}_1$—say that represented by u_0 before the change. We shall obtain the form of the wavefunction at times $t > 0$ and hence the probabilities that a subsequent energy measurement will yield a particular eigenvalue of $\hat{H}_2$.

We first note that the time-dependent Schrödinger equation ensures that a finite discontinuity in $\hat{H}$ produces a similar discontinuity in $\partial\psi/\partial t$, and therefore ψ must be continuous in time. Thus immediately before and after the change we must have

$$\begin{aligned} \Psi(\mathbf{r}, 0) &= u_0(\mathbf{r}) \\ &= \sum_n a_n(0) v_n(\mathbf{r}) \end{aligned} \tag{8.15}$$

where we have used completeness to expand u_0 in terms of the set of eigenfunctions v_n. As $\hat{H}_2$ is time independent we can use (8.4) to obtain an expression for Ψ at all times greater than zero:

$$\Psi(\mathbf{r}, t) = \sum_n a_n(0) v_n(\mathbf{r}) \exp(-i E_n t/\hbar) \tag{8.16}$$

where the energy levels E_n are the eigenvalues of $\hat{H}_2$. Expressions for the constants $a_n(0)$ can be obtained by multiplying both sides of (8.15) by v_n^* and integrating over all space. Thus

$$a_n(0) = \int v_n^* u_0 \, d\tau \tag{8.17}$$

According to the quantum theory of measurement, the probability of obtaining any particular value E_n as a result of a measurement of the energy at any time after the change is equal to $|a_n|^2$. Following such a measurement, of course, the wavefunction would be changed to equal the corresponding eigenfunction v_n.

An example of the practical application of the sudden approximation is the change in the wavefunction of an atom following a radioactive decay of its nucleus. Tritium (^{3}H) can decay by the emission of a β particle and a neutrino to become a positively charged, one-electron ion whose nucleus is ^{3}He. As far as the atomic electron is concerned, therefore, its Hamiltonian has changed suddenly from that corresponding to a hydrogen atom with nuclear charge $Z = 1$ to that of a He^+ ion with $Z = 2$. The energy eigenfunctions of both these systems can be obtained from the expressions given in chapter 3 and the probabilities of

subsequent measurements yielding particular eigenvalues of the He^+ ion can then be readily calculated following the procedure described earlier. For example, the probability of finding the He^+ ion in its ground state is $|A|^2$ where (cf. (8.17))

$$\begin{aligned} A &= \int_0^\infty (8/\pi a_0^3)^{1/2} \exp(-2r/a_0)\ (1/\pi a_0^3)^{1/2} \exp(-r/a_0)\ 4\pi r^2\, dr \\ &= (8\sqrt{2}/a_o^3) \int_0^\infty \exp(-3r/a_0) r^2 dr \\ &= 16\sqrt{2}/27 = 0.838 \end{aligned}$$

so that the probability, $|A|^2$ is 0.70.

A particularly interesting feature of this example follows from the fact that a value of the energy of the He^+ ion can, in principle, be obtained from a knowledge of the energy associated with the nuclear decay, combined with those of the emitted β particle and the neutrino (although in practice the energy of the latter would be very difficult to measure). But the β particle and neutrino could well be a large distance from the atom when these measurements are made so that the energy of the ion would have been measured without apparently interfering with it. Nevertheless, quantum mechanics states that this measurement will cause the wavefunction of the atom to change from a form similar to (8.16) to the appropriate energy eigenfunction. This apparent contradiction is an example of what is known as *entanglement*, which will be discussed more fully in chapters 12 and 13.

8.3 Time-dependent perturbation theory

A very important type of time-dependent problem is one where the Hamiltonian $\hat{H}$ can be written as the sum of a time-independent part $\hat{H}_0$ and a small time-dependent perturbation $\hat{H}'$. An example of this, to which we shall return later, is the case of an atom subject to the oscillating electric field associated with an electromagnetic wave, which can cause transitions to occur from one energy state to another. We shall now describe a method known as time-dependent perturbation theory for obtaining approximate solutions to such problems.

We wish to solve the time-dependent Schrödinger equation (8.1) for the case where

$$\hat{H}(\mathbf{r}, t) = \hat{H}_0(\mathbf{r}) + \hat{H}'(\mathbf{r}, t) \tag{8.18}$$

We assume that the eigenfunctions $u_k(\mathbf{r})$ of $\hat{H}_0$ are known and expand the wavefunction $\Psi(\mathbf{r}, t)$ as a linear combination of these

$$\Psi(\mathbf{r}, t) = \sum_k c_k(t) u_k(\mathbf{r}) e^{-iE_k t/\hbar} \tag{8.19}$$

where the expansion coefficients c_k have been defined so as to exclude the factors $\exp(-iE_k t/\hbar)$, as this simplifies the ensuing argument. Substituting from

(8.18) and (8.19) into (8.1), and assuming that although $\hat{H}'$ is time dependent it commutes with c_k, we get

$$i\hbar \sum_k (\dot{c}_k - i\omega_k c_k) u_k e^{-i\omega_k t} = \sum_k (c_k \hbar \omega_k u_k e^{-i\omega_k t} + c_k \hat{H}' u_k e^{-i\omega_k t})$$

where

$$\dot{c}_k = \frac{\partial c_k}{\partial t} \qquad \text{and} \qquad \omega_k = \frac{E_k}{\hbar}$$

Thus

$$\sum_k (i\hbar \dot{c}_k - c_k \hat{H}') u_k e^{-i\omega_k t} = 0 \tag{8.20}$$

We now multiply (8.20) by the complex conjugate of one of the unperturbed eigenfunctions, u_m^*, and integrate over all space to get

$$i\hbar \dot{c}_m e^{-i\omega_m t} - \sum_k c_k \hat{H}'_{mk} e^{-i\omega_k t} = 0$$

That is

$$\dot{c}_m = \frac{1}{i\hbar} \sum_k c_k \hat{H}'_{mk} e^{i\omega_{mk} t} \tag{8.21}$$

where

$$\hat{H}'_{mk} = \int u_m^* \hat{H}' u_k \, d\tau \qquad \text{and} \qquad \omega_{mk} = \omega_m - \omega_k$$

Everything we have done so far is exact, but we now apply perturbation techniques in a similar manner to that described for the time-independent case in chapter 7. We introduce a constant β, replace $\hat{H}'$ by $\beta \hat{H}'$ and expand the constants c_k in a perturbation series (cf. time-independent perturbation theory, chapter 7)

$$c_k = c_{k0} + \beta c_{k1} \cdots \tag{8.22}$$

Substituting from (8.22) into (8.21), remembering to replace H'_{mk} by $\beta H'_{mk}$ we get

$$\dot{c}_{m0} + \beta \dot{c}_{m1} + \cdots = \frac{1}{i\hbar} \beta \sum_k c_{k0} H'_{mk} e^{i\omega_{mk} t} + \cdots \tag{8.23}$$

where the omitted terms all contain higher-order powers of β. Equating the coefficients of the zeroth and first powers of β we have

$$\dot{c}_{m0} = 0 \tag{8.24}$$

$$\dot{c}_{m1} = \frac{1}{i\hbar} \sum_k c_{k0} H'_{mk} e^{i\omega_{mk} t} \tag{8.25}$$

Equation (8.24) implies that the coefficients c_{m0} are constant in time, which is to be expected as the zero-order Hamiltonian is time independent. The first-order contributions to c_m are obtained from (8.25) as

$$c_{m1} = \frac{1}{i\hbar} \sum_k c_{k0} \int_0^t H'_{mk} e^{i\omega_{mk}t}\, dt \tag{8.26}$$

We are particularly interested in the case where the system is known to be in a particular eigenstate—say that represented by u_n—at the time $t = 0$ so that $c_{n0} = 1$, and $c_{k0} = 0$, $k \neq n$. Equation (8.26) then becomes

$$c_{m1} = \frac{1}{i\hbar} \int_0^t H'_{mn} e^{i\omega_{mn}t}\, dt \tag{8.27}$$

Remembering that $c_{m0} = 0$ ($m \neq n$) we see that the probability of finding the system in a state represented by u_m where $m \neq n$ is given by $|c_{m1}|^2$, provided that c_{m1} is small enough for the perturbation approximation to hold.

Periodic perturbations

We shall now apply these general results to systems where the perturbation varies sinusoidally in time, that is, where

$$\begin{aligned} \hat{H}'(\mathbf{r}, t) &= \hat{H}''(\mathbf{r}) \cos \omega t \\ &= \tfrac{1}{2} \hat{H}''(\mathbf{r}) \big[\exp(i\omega t) + \exp(-i\omega t) \big] \end{aligned} \tag{8.28}$$

and ω is the angular frequency of the perturbation. Substituting into (8.27), we get

$$\begin{aligned} c_{m1} &= \frac{H''_{mn}}{2i\hbar} \int_0^t [e^{i(-\omega+\omega_{mn})t} + e^{i(\omega+\omega_{mn})t}]\, dt \\ &= -\frac{H''_{mn}}{2\hbar} \left[\frac{e^{i(-\omega+\omega_{mn})t} - 1}{-\omega + \omega_{mn}} + \frac{e^{i(\omega+\omega_{mn})t} - 1}{\omega + \omega_{mn}} \right] \end{aligned} \tag{8.29}$$

The first term in square brackets on the right-hand side of (8.29) has a maximum value if $\omega = -\omega_{mn}$ and is comparatively small for values of ω appreciably different from this, while the second term has a similar maximum at $\omega = \omega_{mn}$. It is therefore a reasonable approximation to assume that c_{m1}, and therefore the transition probability, is negligibly small unless one of these conditions is at least approximately, fulfilled. Moreover (except for the special case where $\omega_{mn} = 0$, to which we shall return in the next chapter) both conditions cannot be simultaneously satisfied, so we can assume that only one of the two terms is non-negligible for a particular value of ω. We shall consider the case where $\omega \simeq \omega_{mn}$ and neglect the other term in (8.29) for the moment.

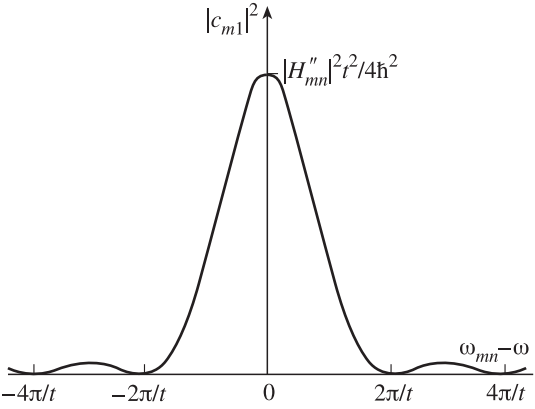


Figure 8.3. The transition probability $|c_{m1}|^2$ as a function of $(\omega_{mn} - \omega)$ in the case of a system which has been subject to a periodic perturbation of angular frequency ω for a time t.

The probability of finding the system in the state u_m at time t is therefore $|c_{m1}|^2$ where

$$|c_{m1}|^2 = \frac{|H''_{mn}|^2}{4\hbar^2}\frac{2[1-\cos(\omega_{mn}-\omega)t]}{(\omega_{mn}-\omega)^2}$$
$$= \frac{|H''_{mn}|^2}{4\hbar^2}\frac{\sin^2[(\omega_{mn}-\omega)t/2]}{[(\omega_{mn}-\omega)/2]^2} \tag{8.30}$$

This expression is plotted as a function of $(\omega_{mn}-\omega)$ in figure 8.3. We note that the height of the central peak is proportional to t^2 while its width is proportional to t^{-1}. Thus, after a time that is long compared with the period of the perturbation, the transition probability will be negligibly small unless the condition $\omega = \omega_{mn}$ is fulfilled. That is, unless

$$\hbar\omega = E_m - E_n \tag{8.31}$$

In the case where the perturbation results from an electromagnetic wave, ω is the angular frequency of the radiation so we see that (8.31) is just the basic equation relating this quantity to the difference between the energies of the states which we discussed in chapter 1. We note that this result has been obtained independently of the photon postulate, which is associated with the quantization of the electromagnetic field. As was emphasized in the discussion of the photoelectric effect in chapter 1, photons are necessary to explain the fact that an atom can be excited into a higher energy state before sufficient classical electromagnetic energy would have reached it. Further evidence for the importance of field quantization comes from the phenomenon of spontaneous emission, which is discussed later.

We can draw another conclusion from (8.31) by noting that, since ω is always positive, E_m must be larger than E_n and the transition discussed corresponds to an excitation of the system, with, presumably, a corresponding absorption of energy from the perturbing field. If, however, we had used the first instead of the second term on the right-hand side of (8.29), we should have obtained the condition $\omega = -\omega_{mn}$ along with expressions similar to (8.30) and (8.31), but with appropriate changes of sign. This would correspond to the case where the perturbation causes a transition from an initial excited state to another of lower energy; energy is then emitted from the system and such a process is known as 'stimulated emission'.

Before we can use (8.30) to calculate more detailed properties of transitions in practical situations, such as the absorption or emission of electromagnetic radiation by atoms in gases, we must allow for the fact that the excited states of quantum systems do not have perfectly defined energies, but are usually broadened into a band containing a large number of closely spaced levels. Furthermore, the perturbation is often not a pure oscillation at a single frequency ω, but is rather a mixture of frequencies in a band of greater or lesser width centred on ω. We first consider the case where the broadening of the energy levels is much greater than that of the perturbation. This broadening can arise from a number of causes: for example the atoms in a gas are in thermal motion so, relative to their own frame of reference, they 'see' the frequency of the perturbation shifted by the Doppler effect: in the laboratory frame of reference this is equivalent to a broadening of the atomic energy levels. Another cause of broadening, to be discussed in more detail shortly, arises because the excited states of atoms always have a 'natural' line width associated with the possibility of spontaneous emission. If one or both of the levels E_m and E_n are broadened for any reason, this results in a similar broadening of the frequency difference ω_{mn}. The latter can then be represented by a function $g(\omega_{mn})$, known as the 'density of states', which is defined so that the number of pairs of levels that have an energy difference between $\hbar\omega_{mn}$ and $\hbar(\omega_{mn} + d\omega_{mn})$ is $g(\omega_{mn})d\omega_{mn}$. Assuming that the matrix elements are identical for all these pairs of states, the total probability $P(t)$ for any such transition to take place is then

$$P(t) = \frac{|H''_{mn}|^2}{4\hbar^2} \int_0^\infty \frac{\sin^2[(\omega_{mn} - \omega)t/2]}{[(\omega_{mn} - \omega)/2]^2} g(\omega_{mn})\, d\omega_{mn} \tag{8.32}$$

For large enough t, $g(\omega_{mn})$ is a more slowly varying function of ω_{mn} than is the rest of the integrand, which is sharply peaked about the point $\omega_{mn} = \omega$ (cf. figure 8.3); we can replace $g(\omega_{mn})$ by $g(\omega)$ and take this quantity outside the integral, which can then be evaluated by standard (though not elementary) methods, leading to

$$P(t) = \frac{\pi |H''_{mn}|^2}{2\hbar^2} g(\omega)t \tag{8.33}$$

The transition rate, W, defined as the transition probability per unit time is then

given by

$$W = \frac{dP}{dt} = \frac{\pi |H''_{mn}|^2}{2\hbar^2} g(\omega) \tag{8.34}$$

Alternatively, the perturbation may not be of a single frequency, but may be made up from a superposition of waves, the number of which having frequencies between ω and $\omega + d\omega$ being $g'(\omega_{mn})$. A similar argument leads to the same expression for W, but with $g(\omega)$ replaced by $g'(\omega_{mn})$. In the general case where both widths have to be taken into account, the relevant quantity is $g''(\omega) = \int g(\omega')g'(\omega - \omega')\, d\omega'$.

Equation (8.34) is known as *Fermi's golden rule*. It has a wide range of applications, both in the field of atomic transitions and, as we shall see in the next chapter, in scattering theory. The important point to remember is that the transition rate per unit time is proportional to the square of the matrix element and to the density of states.

Although energy levels are normally broadened sufficiently for Fermi's golden rule to be applicable, there are some situations in which this is not the case. In a typical laser, for example, the atoms in the lasing medium do not operate independently, but have to be treated as a coherent whole whose transition frequencies are determined by the size and shape of the laser cavity, and the resulting density of states can be very narrow indeed. As a result, the transition probability shown in figure 8.3 does not become narrower than $g(\omega_{mn})$ until t is so large that the value of $|c_{m1}|^2$ at $\omega_{mn} = \omega$ is greater than one, which is of course inconsistent with its definition as a probability. This arises because first-order perturbation theory does not allow for the fact that the probability of occupation of the original state u_n decreases as that relating to u_m rises. A more rigorous analysis, which we shall not describe here, shows that such a system oscillates between the two states at an angular frequency equal to $|H''_{mn}|/2\hbar$. Such 'quantum oscillations' play an important role in the behaviour of lasers and masers, as we shall see in the last section of this chapter where we describe the ammonia maser.

8.4 Selection rules

The results of the previous section will now be used to improve our understanding of the absorption and emission of radiation by atoms undergoing transitions between energy levels. The transition rate clearly determines the intensity of the corresponding spectral line, and both are proportional to the square of the matrix element H''_{mn}. We note, in particular, that if this is zero the transition will not take place. Such non-occurring transitions are described as 'forbidden' and the rules that determine which transitions in a given system are forbidden and which are 'allowed' are known as *selection rules*. These are very important in the study of the physics of atoms, molecules, nuclei and solids; and are extensively discussed in textbooks specializing in these areas. We shall confine our discussion to a brief introduction to the selection rules applying in the case of a one-electron atom.

We consider such an atom subject to a plane electromagnetic wave of angular frequency ω whose electric vector is in the z direction and has magnitude $\mathcal{E}$ where

$$\mathcal{E} = \mathcal{E}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) \tag{8.35}$$

We take the origin of coordinates at the atomic nucleus and assume that the wavelength of the radiation is much greater than the radius of the atom. The term $\mathbf{k} \cdot \mathbf{r}$ is therefore very small so that the electric field is effectively uniform over the atom. The operator representing the energy of interaction between the atom and the field is therefore given by

$$\hat{H}' = e\mathcal{E}z \simeq e\mathcal{E}_0 z \cos \omega t$$

and therefore

$$\hat{H}'' = e\mathcal{E}_0 z \tag{8.36}$$

We now consider the matrix element of this perturbation which connects two states whose quantum numbers are (n_1, l_1, m_1) and (n_2, l_2, m_2) respectively and which we write as H''_{12} for convenience. Using expressions for the one-electron wavefunctions from chapter 3 (3.70), we have

$$\begin{aligned}
H''_{12} &= e\mathcal{E}_0 \int u^*_{n_1 l_1 m_1} z u_{n_2 l_2 m_2} \, d\tau \\
&= e\mathcal{E}_0 \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} R^*_{n_1 l_1}(r) P_{l_1}^{|m_1|}(\cos\theta) e^{-im_1\phi} r \cos\theta \\
&\quad \times R_{n_2 l_2}(r) P_{l_2}^{|m_2|}(\cos\theta) e^{im_2\phi} r^2 \, dr \sin\theta \, d\theta \, d\phi \\
&= e\mathcal{E}_0 \left[\int_0^{\infty} R^*_{n_1 l_1}(r) R_{n_2 l_2}(r) r^3 \, dr \right] \\
&\quad \times \left[\int_0^{\pi} P_{l_1}^{|m_1|}(\cos\theta) P_{l_2}^{|m_2|}(\cos\theta) \cos\theta \sin\theta \, \mathrm{d}\theta \right] \\
&\quad \times \left[\int_0^{2\pi} e^{i(m_2 - m_1)\phi} \, d\phi \right]
\end{aligned} \tag{8.37}$$

It follows that such a transition could take place only if all three integrals in (8.37) are non-zero. If we first consider the integral over ϕ, we see that this vanishes unless $m_1 = m_2$. The integral with respect to θ can be evaluated, given the following property of the associated Legendre functions, $P_l^{|m|}$:

$$(2l+1)\cos\theta P_l^{|m|} = (l - |m| + 1) P_{l+1}^{|m|} + (l + |m|) P_{l-1}^{|m|} \tag{8.38}$$

Using (8.38) and considering the case where $m_1 = m_2$ the integral over θ in (8.37) becomes

$$\frac{l_2 - |m_1| + 1}{2l_2 + 1} \int_0^{\pi} P_{l_1}^{|m_1|} P_{l_2+1}^{|m_1|} \sin\theta \, d\theta + \frac{l_2 + |m_1|}{2l_2 + 1} \int_0^{\pi} P_{l_1}^{|m_1|} P_{l_2-1}^{|m_1|} \sin\theta \, d\theta \tag{8.39}$$

However, associated Legendre functions that have the same value of m but different l are orthogonal, so the integrals in (8.39) vanish unless $l_1 = l_2 + 1$ or $l_1 = l_2 - 1$. So far we have discussed a field polarized in the z direction; if we now consider it to be polarized in the x direction we will have

$$\begin{aligned} \hat{H}'' &= e\mathcal{E}_0 x \\ &= e\mathcal{E}_0 r \sin\theta \cos\phi \end{aligned} \tag{8.40}$$

in spherical polar coordinates. A similar argument to the previous one shows that, in this case, the same relation holds between l_1 and l_2, but the ϕ integral is now

$$\tfrac{1}{2}\int_0^{2\pi} [\exp i(m_2 - m_1 + 1)\phi + \exp i(m_2 - m_1 - 1)\phi] d\phi$$

which is zero unless $m_1 = m_2 \pm 1$. The selection rules governing the allowed transitions in the presence of an electromagnetic wave of arbitrary polarization are therefore

$$\left.\begin{aligned} \Delta l &= \pm 1 \\ \Delta m &= \pm 1 \text{ or } 0 \end{aligned}\right\} \tag{8.41}$$

where Δl and Δm represent the differences between the values of l and m associated with the two states. We note that there are no such rules governing the change in the principal quantum number n, which results from the fact that the first integral in (8.37) is always finite; its value can, however, vary considerably from one transition to another resulting in a corresponding variation in the intensities of the spectral lines.

These selection rules refer to *electric dipole transitions* where the perturbing Hamiltonian has the form (8.36) or (8.40). However, other types of transition can be induced by electromagnetic radiation. First, $(\mathbf{k}\cdot\mathbf{r})$ can be assumed to be small rather than zero, and the right-hand side of (8.35) can be expanded as a power series leading to

$$\mathcal{E} = \mathcal{E}_0[\cos\omega t + (\mathbf{k}\cdot\mathbf{r})\sin\omega t - \tfrac{1}{2}(\mathbf{k}\cdot\mathbf{r})^2\cos\omega t + \cdots] \tag{8.42}$$

Only the first term of (8.42) has been considered so far and the others can lead to transitions known as *electric quadrupole, electric octopole*, etc. Second, the atomic electron interacts with the magnetic field associated with the wave as well as the electric field. This leads to the possibility of transitions known as *magnetic dipole, magnetic quadrupole*, etc. Detailed calculations of the transition probabilities associated with both these classes of transition are most conveniently performed if the electromagnetic wave is expressed in vector-potential form. Details of this and the resulting selection rules can be found in textbooks on atomic physics: in general the spectral lines associated with allowed electric dipole transitions are much more intense than any others.

Spontaneous emission

So far we have only discussed the case where transitions between energy levels are caused by the perturbation associated with an applied field. However, it is well known that an atom in an excited state will decay to its ground state, emitting radiation, even if it is completely isolated. This appears to contradict our previous statement that a system in an energy eigenstate should remain in this state indefinitely, and to resolve this we have to consider the quantization of the electromagnetic field. The Hamiltonian (8.28) is evaluated assuming the classical form of the electromagnetic wave, neglecting the fact that it is quantized and consists of photons. A proper discussion of field quantization is well outside the scope of this book, although we touch on it again in chapter 11, but the following simplified argument explains its application to the present problem.

Classically, an electromagnetic wave consists of perpendicular electric and magnetic fields and we first consider a single mode of oscillation with angular frequency ω. The energy of such a mode can be written as

$$W = \frac{1}{2}\varepsilon_0\langle\mathcal{E}^2\rangle + \frac{1}{2\mu_0}\langle B^2\rangle$$

per unit volume, where $\langle\mathcal{E}^2\rangle$ and $\langle B^2\rangle$ are the mean square amplitudes of the electric and magnetic fields. This expression bears some similarity to that for the energy of a classical harmonic oscillator of frequency ω_c:

$$E = \frac{p^2}{2m} + \frac{1}{2}m\omega_c^2 x^2.$$

It turns out that this analogy can be developed to quantize the electromagnetic field. Just as a harmonic oscillator has a zero-point energy $\frac{1}{2}\hbar\omega_c$ (cf. chapter 2, section 2.6) a mode of the electromagnetic field with frequency ω has a minimum possible energy of $\frac{1}{2}\hbar\omega$. This means that even in the vacuum, where classically we would not expect any fields to exist, each possible mode of oscillation has this energy. The vacuum must therefore contain fluctuating electric and magnetic fields, and the mean square amplitude of the former is given by

$$\langle\mathcal{E}^2\rangle = \hbar\omega/(2\varepsilon_0 V) \tag{8.43}$$

where V is the volume occupied by the field. This field can cause an excited atom to 'spontaneously' decay with the emission of a photon. The reverse process is clearly impossible as energy must be conserved and the field energy is already at its minimum value. We can use this, along with our earlier results, to estimate the expected rate of spontaneous emission from an excited state. To do this we need to know the density of states, or number of modes of oscillation, associated with the field. This is equal to the number of modes of vibration in a volume V, which can be calculated by assuming the volume to be a square box of side L on the

boundaries of which the electric field must be zero. It follows that the **k** vectors of the waves must satisfy the condition

$$\mathbf{k} = (k_x, k_y, k_z) = (n_1\pi/L, n_2\pi/L, n_3\pi/L) \tag{8.44}$$

where n_1, n_2 and n_3 are integers. There are therefore two modes (with different polarizations) in the volume π^3/V in 'k-space'. Waves whose wavevectors have magnitudes in the range k to $k + dk$ have angular frequencies between ck and $c(k + dk)$. These can point in any direction in the positive octant of k-space so the total number with angular frequency between ω and $\omega + d\omega$ is

$$2 \times (\tfrac{1}{2}\pi\omega^2\, d\omega/c^3)(V/\pi^3) = (V\omega^2/2\pi^2c^3)\, d\omega \tag{8.45}$$

This must equal $g'(\omega)\, d\omega$, where g' is the density of states, so we can combine this with (8.34) and (8.35) to calculate the transition rate as

$$W_{\text{sp}} = (e^2\omega^3/8\pi\varepsilon_0\hbar c^3)|z_{12}|^2 \tag{8.46}$$

We note that this rate is proportional to $|z_{12}|^2$ so spontaneous emission is subject to the same selection rules as applied to stimulated emission and absorption discussed earlier.

8.5 The Ehrenfest theorem

The time-dependent Schrödinger equation can be used to calculate the rate of change of the expectation value of a physical quantity

$$\begin{aligned}\frac{\partial\langle\hat{Q}\rangle}{\partial t} &= \frac{\partial}{\partial t}\int \Psi^*\hat{Q}\Psi\, d\tau \\ &= \int\left[\frac{\partial\Psi^*}{\partial t}\hat{Q}\Psi + \Psi^*\hat{Q}\frac{\partial\Psi}{\partial t} + \Psi^*\frac{\partial\hat{Q}}{\partial t}\Psi\right] d\tau\end{aligned}$$

We can now substitute expressions for $\partial\Psi/\partial t$ and $\partial\Psi^*/\partial t$ obtained from the time-dependent Schrödinger equation (8.1) and its complex conjugate to get

$$\begin{aligned}\frac{\partial\langle\hat{Q}\rangle}{\partial t} &= \int\frac{i}{\hbar}[(\hat{Q}\Psi)\hat{H}^*\Psi^* - \Psi^*\hat{Q}\hat{H}\Psi]\, d\tau + \left\langle\frac{\partial\hat{Q}}{\partial t}\right\rangle \\ &= \frac{i}{\hbar}\int[\Psi^*(\hat{H}\hat{Q} - \hat{Q}\hat{H})\Psi]\, d\tau + \left\langle\frac{\partial\hat{Q}}{\partial t}\right\rangle\end{aligned}$$

using the Hermitian property of $\hat{H}$. That is

$$\frac{\partial\langle\hat{Q}\rangle}{\partial t} = \frac{i}{\hbar}\langle[\hat{H}, \hat{Q}]\rangle + \left\langle\frac{\partial\hat{Q}}{\partial t}\right\rangle \tag{8.47}$$

where $[\hat{H}, \hat{Q}]$ is the commutator of $\hat{Q}$ and $\hat{H}$. Thus the rate of change of the expectation value of the physical quantity represented by $\hat{Q}$ is equal to a sum of two terms: the first is proportional to the expectation value of the commutator of $\hat{Q}$ with the Hamiltonian of the system, $\hat{H}$, while the second is equal to the expectation value of $\partial\hat{Q}/\partial t$. This result is known as the Ehrenfest theorem. We note that in the particular case where $\hat{Q}$ is time independent and commutes with $\hat{H}$, the expectation value of $\hat{Q}$ is constant, in agreement with the fact that the value of the corresponding physical quantity is conserved (cf. section 4.6).

We can use the Ehrenfest theorem to investigate further the connection between quantum and classical mechanics. In general we find that the expectation values of the quantum-mechanical quantities are related by the classical equations of motion. For example, the fact that energy is conserved in a closed system follows by putting $\hat{Q}$ equal to $\hat{H}$, remembering that in such a case the latter is time independent. Alternatively, if we put $\hat{Q}$ equal to $\hat{x}$, (8.47) gives

$$\frac{\partial\langle\hat{x}\rangle}{\partial t} = \frac{i}{\hbar}\left\langle\left[\left(\frac{\hat{P}_x^2}{2m} + V\right), \hat{X}\right]\right\rangle$$

We know that $\hat{X} \equiv x$ commutes with $V(x)$ and that $[\hat{P}_x, x] = -i\hbar$, so that

$$\begin{aligned}[\hat{P}_x^2, x] &= \hat{P}_x^2 x - x\hat{P}_x^2 \\ &= \hat{P}_x(-i\hbar + x\hat{P}_x) - (i\hbar + \hat{P}_x x)\hat{P}_x \\ &= -2i\hbar\hat{P}_x\end{aligned}$$

so that

$$\frac{\partial\langle\hat{x}\rangle}{\partial t} = \left\langle\frac{\hat{P}_x}{m}\right\rangle \tag{8.48}$$

which corresponds to the classical definition of momentum. Similarly if $\hat{Q} = \hat{P}_x$,

$$\frac{\partial\langle\hat{P}_x\rangle}{\partial t} = \frac{i}{\hbar}[V(x), P_x] = -\left\langle\frac{\partial V}{\partial x}\right\rangle \tag{8.49}$$

where we have used the relation $\hat{P}_x = -i\hbar\partial/\partial x$. Equation (8.49) simply states that the rate of change in the expectation value of the x component of momentum is equal to the expectation value of the x component of the applied force, which of course corresponds to Newton's second law.

Thus the correspondence principle follows from the Ehrenfest theorem, provided we can identify the quantum-mechanical expectation value of a physical quantity with its value in classical mechanics. In most classical systems this condition is clearly satisfied: macroscopic particles are normally strongly localized and wavefunction spreads are therefore very narrow compared with the dimensions of the measuring apparatus.

However, there are some circumstances in which the results of quantum mechanics do not go over to those of classical mechanics in such a direct way. Consider, for example, the case of a particle passing through the two slits of a Young's interference experiment; its wavefunction will be finite over the area of each slit, but the expectation value of its position corresponds to a point midway between the slits. At this point the wavefunction is zero and the particle is never observed at a position corresponding to this expectation value—however far apart the slits or however massive the particle. What happens in this case is that the interference effects associated with the delocalization of the particle become harder and harder to observe in the classical limit and the system becomes indistinguishable from one in which the particle passes through either one of the two slits, but not both. Nevertheless, there is a theoretical possibility that a macroscopic system could be constructed which displayed quantum-mechanical delocalization and whose properties could not be accounted for classically. Some of the conceptual problems associated with such macroscopic quantum effects are discussed in chapter 13.

8.6 The ammonia maser

A physical system that illustrates a number of the quantum-mechanical results discussed in this and previous chapters is the ammonia maser. A maser is a device that amplifies microwave radiation. It relies on the provision of a medium (in this case ammonia) that has more of its molecules in an excited state than there are in the ground state. Radiation of the correct frequency then induces transitions to the ground state and the consequent stimulated emission of more radiation. The way in which this process operates in the case of ammonia will be described in this section.

The ammonia molecule consists of a nitrogen atom bound chemically to three hydrogen atoms in such a way that the nitrogen atom lies a little above (or below) the centre of the equilateral triangle formed by the hydrogens (see figure 8.4). The exact wavefunction describing the four nuclei and ten electrons in the ammonia molecule is very complicated and would require considerable computational effort to evaluate—if indeed this were even possible. However, all we need to know for our present discussion is that there are two states which are equivalent apart from the position (up or down) of the nitrogen nucleus (see figure 8.4) and the fact that it is possible for the atom to pass from one position to the other by quantum-mechanical tunnelling.[1]

However, neither of the two states described so far is an energy eigenstate. This follows from the fact that the molecule is able to tunnel between the two configurations and that, as the potential is centrosymmetric, the energy eigenstates

[1] It might be thought that the transition from the 'up state' to the 'down state' could be effected simply by rotating the molecule about an axis in the plane of the hydrogens. In fact in the relevant states the molecule possesses angular momentum about its symmetry axis which stabilizes its spatial orientation gyroscopically.

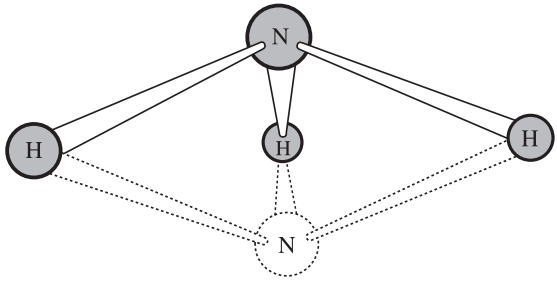


Figure 8.4. The geometrical configuration of the ammonia molecule in states where the nitrogen atom is above (continuous lines) and below (broken lines) the plane of the three hydrogens.

must have definite parity—that is, they must be either symmetric or antisymmetric with respect to reflection. If we call the wavefunctions corresponding to the nitrogen in the up and down positions ϕ_1 and ϕ_2 respectively, we can make linear combinations (u_1 and u_2) of these which do have definite parity and which are then reasonable approximations to the energy eigenfunctions:

$$\begin{aligned} u_1 &= \frac{1}{\sqrt{2}}(\phi_1 + \phi_2) \\ u_2 &= \frac{1}{\sqrt{2}}(\phi_1 - \phi_2) \end{aligned} \tag{8.50}$$

It follows that, if we know the position of the nitrogen atom relative to the hydrogens so fixing the wavefunction as ϕ_1 or ϕ_2, the molecule cannot be in a state of definite energy. However, if we know the energy, the nitrogen position must be uncertain. We note in passing that, although a nitrogen atom is not macroscopic, it is a lot heavier than an electron, and the fact that it can be delocalized is an interesting confirmation of the application of quantum mechanics to such an object.

The eigenfunction u_1 is symmetric and corresponds to the ground state of the system whose energy we represent by E_1, while u_2 is antisymmetric and is therefore an excited state (energy E_2). It turns out that the energy difference corresponds to a frequency in the microwave region and we should therefore be able to construct a maser provided that we can produce nitrogen molecules entirely, or at least predominantly, in the state u_2. Now, at normal temperatures, the average thermal energy of the molecules is much greater than the difference in energy between the two states, so if their energy were to be measured, an approximately equal number would be found in each state. In order to isolate a set of molecules all in the state u_2, we first create a molecular beam by allowing the gas to emerge from a container through a fine hole (see figure 8.5). We then measure the energy of each molecule, so placing it in one of the energy

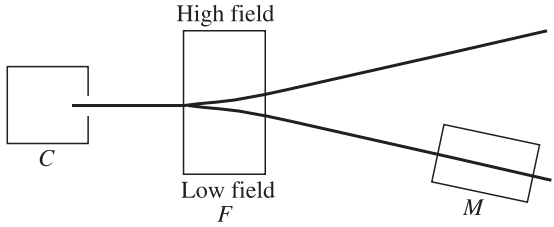


Figure 8.5. The ammonia maser. Ammonia molecules leave the container C and pass through a region of non-uniform electric field $\mathcal{E}$. Excited molecules are deviated towards the lower field region and directed into a microwave cavity M.

eigenstates. This is achieved by passing the beam through a region where there is a static non-uniform electric field, the effect of which we will now explain by considering the electrical polarizability of the ammonia molecule.

A general expression for the polarizability of a one-electron atom was obtained by perturbation theory in chapter 7 (equation (7.29)). This can be applied to an ammonia molecule, provided the operator ez representing the instantaneous electric dipole moment of the atom is replaced by the corresponding molecular property. It is known that the electrons in the ammonia molecule are distributed in such a way that the nitrogen atom carries a net negative charge, which we denote by $-q$, while a positive charge of equal magnitude is distributed between the three hydrogens. It follows that, if the nitrogen atom were at a distance x from the plane containing the hydrogens the molecule would have a dipole moment qx, and this quantity corresponds with ez in the atomic case. Moreover, the difference between the energies of the states u_1 and u_2 is much less than that between either of them and any other energy state of the system. These latter ones can therefore be ignored when considering the perturbation expression for the polarizabilities of the states which are therefore

$$\left.\begin{aligned}\alpha_1 &= \frac{2q^2|x_{12}|^2}{E_2 - E_1}\\ \alpha_2 &= \frac{2q^2|x_{12}|^2}{E_1 - E_2}\end{aligned}\right\} \tag{8.51}$$

where $x_{12} = \int u_1^* x u_2 \, d\tau$.[2] It follows that α_1 is positive while α_2 is negative, and the application of an electric field will induce a net dipole moment parallel to the field in the first case and antiparallel in the second. The application of a non-uniform field therefore sorts the molecules into two groups: one, consisting of molecules in their ground state, moves to a region of high field intensity, while the other, containing only molecules in the excited state, moves to low-field regions.

[2] Of course the field must be weak enough for the polarization energy to be much less than $E_2 - E_1$, otherwise the two energy eigenstates would be mixed as in the Stark effect.

This process is analogous to the measurement of a spin component using a non-uniform magnetic field in the Stern–Gerlach experiment.

Having separated out a beam of excited molecules, the next step is to direct it into a microwave cavity whose dimensions are such that its natural frequency equals that associated with the transition. Microwave radiation of this frequency which enters the cavity will therefore stimulate the emission of more radiation, resulting in the required amplification. In fact, the transition probability is typically so large and the line width in the cavity is so small, that quantum oscillations of the type described earlier in this chapter can occur: in one half-cycle of such an oscillation when the molecules are predominantly excited, amplification results, but in the other when the molecules are largely in the ground state, the incident radiation is absorbed. To avoid this, the speed of the molecular beam is chosen so that each molecule is in the cavity for a time approximately equal to half the period of the quantum oscillation; since on entry it is in its excited state, the absorption phase is thereby eliminated and continuous amplification results.

Problems

8.1 The wavefunction of a particle in a one-dimensional infinite-sided potential well of width $2a$ is $\psi = 2^{-1/2}(u_1 + u_2)$ at time $t = 0$ where u_1 and u_2 are the two lowest energy eigenfunctions. Find an expression for the position probability distribution as a function of time and show that it is periodic with angular frequency $\omega = 3\pi^2\hbar/8ma^2$. Sketch this probability distribution at times 0, $\pi/2\omega$, π/ω and $3\pi/2\omega$.

8.2 Show that the expectation value of the position of a particle in a harmonic oscillator potential oscillates sinusoidally with the classical frequency if the system is not in an energy eigenstate.

8.3 Show that the expectation value of the angular momentum of an electron in a magnetic field **B** precesses about the direction of **B** with an angular frequency eB/m_e, unless it is in an energy eigenstate.

8.4 What is the probability of finding the resulting $^3He^+$ ion in (i) its $2s$ and (ii) one of its $2p$ states, following the β decay of a ^3H atom initially in its ground state?

8.5 The spring constant of a harmonic oscillator in its ground state is suddenly doubled. Calculate the probability that a subsequent energy measurement will find the new oscillator in (i) its ground state, (ii) its first excited state, and (iii) its second excited state.

8.6 In the experiment on spin interference described in section 8.1, the magnetic fields used had a magnitude of $0.5T$ and the path lengths were each 7×10^{-5} m. Show that a minimum in the centre of the diffraction pattern is to be expected for neutrons with a wavelength of 3.89×10^{-10} m, given that the neutron has a magnetic moment of magnitude 1.91 $(e\hbar/2m_n)$, m_n being the neutron mass.

8.7 A particle, initially in an energy eigenstate of an infinite-sided potential well, is subject to a perturbation of the form $V_0 x \cos\omega t$. Show that transitions are possible between the states u_n and u_m only if $n + m$ is odd.

8.8 The amplitude H'' associated with magnetic dipole transitions turns out to be proportional to the operator representing the angular-momentum vector. Show that in a one-electron atom the selection rules for such transitions are $\Delta l = 0$, $\Delta m = \pm 1$ or 0. Which of these apply when the angular-momentum vector is (i) parallel and (ii) perpendicular to the z axis?

Hint: Express L_x and L_y in terms of ladder operators.

8.9 Given that $yH_n = \frac{1}{2}H_{n+1} + nH_{n-1}$ where the H_n's are Hermite polynomials, show that the selection rule for electric dipole transitions in a one-dimensional harmonic oscillator is $\Delta n = \pm 1$.

8.10 A certain physical system has a Hamiltonian operator of the form $\hat{H}_0 + \hat{H}'' \cos \omega t$ where $\hat{H}_0$ and $\hat{H}''$ are time independent, but H'' need not be a small perturbation. The operator $\hat{H}_0$ has only two eigenstates whose eigenfunctions are u_1 and u_2 respectively. Show that the expression $au_1 \exp(-iE_1 t/\hbar) + bu_2 \exp(-iE_2 t/\hbar)$ is a solution to the time-dependent Schrödinger equation in this case if

$$\frac{da}{dt} = \frac{b}{i\hbar} H''_{21} \cos \omega t e^{i\omega_{21} t}$$

and

$$\frac{db}{dt} = \frac{a}{i\hbar} H''_{21} \cos \omega t e^{i\omega_{12} t}$$

in the usual notation, provided $H''_{11} = H''_{22} = 0$. Show that if $\omega = \omega_{12}$ and if high-frequency terms can be ignored, then

$$a = \cos(\Omega t - \phi) \qquad b = \sin(\Omega t - \phi)$$

are solutions to these equations where $\Omega = |H''_{12}|/2\hbar$ and ϕ is a constant. Compare these results with the discussion of quantum oscillations in the text.

8.11 A particle moves in a one-dimensional potential well given by

$$V = V_0 \quad (-a \leqslant x \leqslant a) \qquad V = 0 \quad (a < |x| < b) \qquad V = \infty \quad (|x| > b)$$

Assuming that the magnitudes of a, b, and V_0 are such that the ground-state energy is less than V_0 and that there is a small, but finite, probability of tunnelling through the central barrier, draw sketch graphs of the two lowest energy eigenfunctions of this system. Discuss the evolution in time of such a system if the particle is known to be initially on one side of the barrier. Consider the response of this system to a perturbation whose frequency corresponds to the difference between the energies of the two lowest states and compare the properties of this system with those of the ammonia molecule.

Chapter 9

Scattering

In this chapter we discuss the quantum mechanics of the scattering of particles by a fixed object. Scattering is a very important feature of many physical processes and is nearly always a part of the process of performing and recording an experimental result. Thus when we perform an x-ray diffraction experiment we record the diffraction pattern created when x-rays are scattered by a crystal, while nearly all the information we possess about the energy levels and structure of nuclei and fundamental particles has been obtained from experiments in which beams of particles (e.g. protons or neutrons) are scattered from targets containing the nuclei or particles under investigation. Indeed any visual observation we make simply by looking at something involves the detection by our eyes of light scattered from the object we are looking at. It is clearly important therefore that we understand something of the quantum mechanics of the scattering processes, and this will be the aim of the present chapter.

We shall begin our study of the quantum mechanics of scattering by a consideration of a simple one-dimensional example which will illustrate a number of the points to be taken up in the subsequent discussion of three-dimensional systems. We shall introduce the latter by some general considerations; this will be followed by a description of the Born approximation in which the scattering process is treated as a perturbation, and we shall finally introduce the method known as partial wave analysis, which is particularly useful in the description of the scattering of plane waves by spherical objects. Throughout this chapter we shall assume that the particles being scattered are of a different type from the object or particle doing the scattering, and we shall briefly discuss the particular features that arise in the case of collisions between identical particles towards the end of chapter 10.

9.1 Scattering in one dimension

Scattering experiments are usually performed using beams of particles all initially moving with the same speed in the same direction, so we begin our study by

considering the wavefunction of a particle moving with momentum p in the positive x direction. We know this must be an eigenfunction of the momentum operator $\hat{P}_x = -i\hbar\partial/\partial x$ so that

$$\psi_k(x) = A_k \exp(ikx) \tag{9.1}$$

where the momentum eigenvalue is $\hbar k$. Wavefunctions corresponding to different values of k should be orthogonal and the constants A_k should be chosen to ensure that the wavefunctions are normalized. However, functions of the form (9.1) extend throughout the one-dimensional space and the usual normalization procedure would imply that $A = 0$—because, if the particle can be anywhere between plus and minus infinity with equal probability, the probability of finding it in the vicinity of any particular point should be zero. The difficulty is associated with the fact that the eigenvalues form a continuous set, and in chapter 4 we discussed how such problems could be treated using delta-function normalization. We now take a different approach which consists of redefining ψ_k so as to represent a *beam* containing many particles rather than a single particle, and we assume that the average separation of the particles in the beam is L. If we now normalize ψ_k so that $|\psi_k(x)|^2\,dx$ represents the probability of finding *any* particle in the region between x and $x + dx$ we get

$$\int_0^L \psi_k^* \psi_k \, dx = 1 \tag{9.2}$$

and hence, substituting from (9.1)

$$\psi_k = L^{-1/2} \exp(ikx) \tag{9.3}$$

It would clearly be very convenient if the orthogonality condition could be expressed in a similar manner so that

$$\int_0^L \psi_k^* \psi_{k'} \, dx = \delta_{kk'} \tag{9.4}$$

Substitution of (9.3) into the left-hand side of (9.4) shows that the latter equation can be valid only if the values of k are restricted so that

$$k = 2n\pi/L \tag{9.5}$$

where n is an integer. We note that this condition is identical to that derived in our discussion of the one-dimensional solid in chapter 7 where we imposed 'periodic boundary conditions' so that $\psi(x + L) = \psi(x)$ and we shall assume these to be valid here also. As we discussed in chapter 7, no physical significance should be attached to the periodic boundary conditions: they are a mathematical device which enables us to impose orthonormality on free-particle wavefunctions and hence to calculate properties such as the density of states (see later).

Particle flux

A property of particle beams that will be particularly useful in the subsequent discussion is known as particle flux. In one dimension this is defined as the average number of particles passing a point per unit time and is represented by Γ. If the particles are in a momentum eigenstate, Γ equals the number of particles per unit length multiplied by the particle velocity: that is,

$$\Gamma = \hbar k/mL \tag{9.6}$$

where m is the particle mass. We now obtain an expression for the flux in the general case where the one-dimensional wavefunction is $\Psi(x, t)$.

Consider a region of the x axis between the points $x = x_1$ and $x = x_2$ where $x_2 > x_1$. The probability of finding a particle in this region is P where

$$P = \int_{x_1}^{x_2} \psi^* \psi \, dx \tag{9.7}$$

Assuming that particles cannot be created or destroyed, the net flux of particles into this region must be equal to the rate of change in time of P. It follows that

$$\Gamma(x_1) - \Gamma(x_2) = \frac{\partial P}{\partial t} = \int_{x_1}^{x_2} \left(\Psi^* \frac{\partial \Psi}{\partial t} + \Psi \frac{\partial \Psi^*}{\partial t} \right) dx \tag{9.8}$$

We can now substitute for $\partial\Psi/\partial t$ and $\partial\Psi^*/\partial t$ from the time-dependent Schrödinger equation (8.1) to get

$$\Gamma(x_1) - \Gamma(x_2) = \frac{i\hbar}{2m} \int_{x_1}^{x_2} \left(\Psi^* \frac{\partial^2 \Psi}{\partial x^2} - \Psi \frac{\partial^2 \Psi^*}{\partial x^2} \right) dx \tag{9.9}$$

where terms involving the potential energy have cancelled each other out. We now integrate by parts to obtain

$$\Gamma(x_1) - \Gamma(x_2) = \frac{i\hbar}{2m} \left[\Psi^* \frac{\partial \Psi}{\partial x} - \Psi \frac{\partial \Psi^*}{\partial x} \right]_{x_1}^{x_2} \tag{9.10}$$

Equation (9.10) is valid for all pairs of points x_1 and x_2, so the flux past the point x must be given by

$$\Gamma(x) = -\frac{i\hbar}{2m} \left(\Psi^* \frac{\partial \Psi}{\partial x} - \Psi \frac{\partial \Psi^*}{\partial x} \right) \tag{9.11}$$

We note that, if we substitute the momentum eigenfunction (9.3) into (9.11), we get the same expression (9.6) for the particle flux as that obtained earlier.

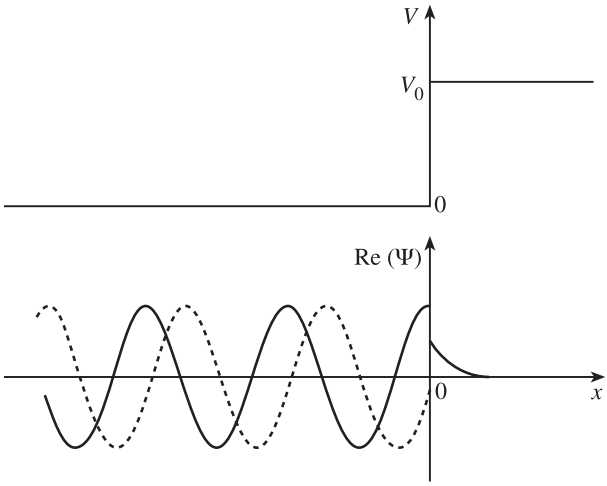


Figure 9.1. The potential V and the real parts of the incident, reflected, and transmitted waves in the case of the scattering of particles by a potential step.

Scattering by a potential step

We now consider the case where a beam of particles is scattered by a simple potential step of the form illustrated in figure 9.1. That is, we are considering the motion of particles in a potential $V(x)$ where

$$\left.\begin{aligned} V(x) &= 0 \qquad x \leqslant 0 \\ V(x) &= V_0 \qquad x > 0 \end{aligned}\right\} \tag{9.12}$$

We are interested in the steady-state solution to the problem, ignoring any transient effects associated with the 'switching on' of the beam, which means that, as the potential (9.12) is time independent, the wavefunction should be one of the energy eigenfunctions of the system. Now in the region where the potential is zero the momentum eigenfunctions with eigenvalues $\pm\hbar k$ are also degenerate eigenfunctions of the energy with energy $\hbar^2k^2/2m$. A general expression for the energy eigenfunction in the region $x < 0$ is therefore

$$u(x) = L^{-1/2}(e^{ikx} + \alpha e^{-ikx}) \tag{9.13}$$

where α is a constant. In the study of scattering we try to predict the results of experiments that measure the flux of the incident and scattered wavefunctions separately. In one dimension this means that we have one detector in the region $x < 0$ which identifies and counts particles moving in (say) the positive direction while leaving the others undisturbed, and another which performs a similar measurement on the particles moving in the negative direction. The energy

eigenfunction (9.13) then represents the wavefunction *before* these measurements are performed and the probabilities of different results can be predicted using the standard procedure set out in the basic postulates. Thus, in the present case (9.13) tells us that a measurement of the flux of particles moving in the positive k direction will produce a result $\hbar k/mL$ while that in the negative direction is $|\alpha|^2\hbar k/mL$, so that the probability of a particular particle being scattered is $|\alpha|^2$.

To obtain an expression for the constant α we must consider the wavefunction in the region $x > 0$ where its form depends on whether the incident kinetic energy $E = \hbar^2k^2/2m$ is greater or less than V_0. We shall consider the latter case first when the solution to the time-independent Schrödinger equation is easily seen to be

$$u(x) = C\exp(-\kappa x) \qquad x > 0 \tag{9.14}$$

where $\kappa = [2m(V_0 - E)/\hbar^2]^{1/2}$ and we have rejected the corresponding solution with a positive exponent because it implies a divergence of u as x tends to infinity, in breach of the boundary conditions applying to one-dimensional systems.

We can now apply the conditions that the wavefunction and its first spatial derivative be continuous at $x = 0$ (cf. chapter 2) to get

$$L^{-1/2}(1 + \alpha) = C$$
$$L^{-1/2}ik(1 - \alpha) = -\kappa C$$

and hence

$$\left.\begin{aligned}\alpha &= -(\kappa + ik)/(\kappa - ik)\\ C &= -[2ik/(\kappa - ik)]L^{-1/2}\end{aligned}\right\} \tag{9.15}$$

The scattering probability $|\alpha|^2$ is therefore seen to be unity so that all the particles incident on the potential step are scattered and none are transmitted. This point can be confirmed by substituting from (9.14) into (9.11) which produces a zero value for the transmitted flux. The net result of the scattering is therefore to produce a scattered wave whose amplitude is the same as that of the incident wave. However, its phase at the point $x = 0$ is increased relative to the phase of the incident wave by an amount δ, where δ is the phase of α. Using (9.15), we get

$$\delta = \tan^{-1}[2k\kappa/(\kappa^2 - k^2)]$$

Clearly a similar result to this will apply in any one-dimensional scattering problem where the potential barrier is high enough to prevent the particles from being transmitted, even if it does not have the simple step form. The total wavefunction in a region on the incident side of the scatterer where the potential is zero will always consist of a sum of incident and reflected plane waves of equal amplitude, and the detailed nature of the scattering potential will affect only the phase shift δ. Similar phase shifts turn out to be of considerable importance in three-dimensional scattering and we shall return to this concept later in the chapter when we discuss the method of 'partial waves'.

We now turn to the case where the kinetic energy of the particles is greater than the height of the barrier so that we expect there to be a finite probability of the particle being transmitted. The wavefunction in the region $x < 0$ still has the general form (9.13) while the transmitted wave is clearly

$$u(x) = C' \exp(ik'x) \tag{9.16}$$

where

$$k' = [2m(E - V_0)/\hbar^2]^{1/2}$$

and C' is obtained along with α by applying the continuity conditions in the same way as before, leading to

$$\left.\begin{aligned} \alpha &= \frac{k - k'}{k + k'} \\ C' &= \frac{2k}{k + k'} L^{-1/2} \end{aligned}\right\} \tag{9.17}$$

The reflected flux is therefore

$$\frac{\hbar k}{mL}|\alpha|^2 = \frac{\hbar k}{mL}\frac{(k - k')^2}{(k + k')^2}$$

while the transmitted flux is

$$\frac{\hbar k'}{m}|C'|^2 = \frac{\hbar k}{mL}\frac{4kk'}{(k + k')^2}$$

It follows directly that the scattering and transmission probabilities are

$$\frac{(k - k')^2}{(k + k')^2} \qquad \text{and} \qquad \frac{4kk'}{(k + k')^2}$$

respectively and we note that the sum of these two probabilities is equal to one as expected.

9.2 Scattering in three dimensions

We turn to the three-dimensional case and begin with some general statements about particle flux and scattering probability. Whereas in one dimension the particle flux was defined as the number of particles passing a point per second, in three dimensions we define *flux density* as a vector quantity $\mathbf{\Gamma}$ such that $\mathbf{\Gamma} \cdot d\mathbf{A}$ is the total flux of particles passing through the element of area $d\mathbf{A}$ per second. Clearly the direction of $\mathbf{\Gamma}$ corresponds to the direction of motion of the particles at the point under consideration, while its magnitude represents the number of particles crossing unit area per second. We shall now obtain an expression for $\mathbf{\Gamma}$

in the case of a system represented by the wavefunction $\Psi(\mathbf{r}, t)$ using a procedure similar to that employed in the one-dimensional case.

Consider a volume V enclosed by a surface A. The net number of particles entering V through A in unit time must be equal to the rate of increase in the probability of finding a particle in V. That is

$$-\oint_A \mathbf{\Gamma}(\mathbf{r}) \cdot d\mathbf{A} = \frac{\partial}{\partial t} \int_V \Psi^* \Psi \, d\tau$$
$$= \int_V \left(\frac{\partial \Psi^*}{\partial t} \Psi + \Psi^* \frac{\partial \Psi}{\partial t} \right) d\tau \tag{9.18}$$

Following a procedure similar to that applied in the one-dimensional case—(9.9)–(9.11)—we substitute for $(\partial\Psi/\partial t)$ and $(\partial\Psi^*/\partial t)$ from the time-dependent Schrödinger equation, the potential energy terms cancel out as before and we get

$$\oint_A \Gamma(\mathbf{r}) \cdot d\mathbf{A} = -\frac{i\hbar}{2m} \int_V (\Psi^* \nabla^2 \Psi - \Psi \nabla^2 \Psi^*) \, d\tau$$
$$= -\frac{i\hbar}{2m} \int_A (\Psi^* \nabla \Psi - \Psi \nabla \Psi^*) \cdot d\mathbf{A} \tag{9.19}$$

where we have applied the theorem in vector calculus known as Green's theorem which corresponds to integration by parts in three dimensions. The expression for $\mathbf{\Gamma}(\mathbf{r})$ follows directly, remembering that (9.19) is valid for any closed surface:

$$\mathbf{\Gamma}(\mathbf{r}) = -\frac{i\hbar}{2m} (\Psi^* \nabla \Psi - \Psi \nabla \Psi^*) \tag{9.20}$$

which is therefore the three-dimensional equivalent of (9.11).

As an example we consider the special case of a beam of free particles of momentum $\hbar\mathbf{k}$ whose wavefunction is

$$\Psi(\mathbf{r}, t) = V^{-1/2} \exp i(\mathbf{k} \cdot \mathbf{r} - Et/\hbar) \tag{9.21}$$

where $E = \hbar^2 k^2 / 2m$ and the beam has been normalized so that it contains, on average, one particle in a volume V (which is, of course, not necessarily the same as the volume V used previously). Substituting from (9.21) into (9.20) we have

$$\Gamma = \frac{\hbar \mathbf{k}}{mV} \tag{9.22}$$

which is the same expression as would be obtained from elementary considerations remembering that in this case all particles have velocity $\hbar\mathbf{k}/m$.

Scattering cross section

If a beam of particles is incident on a scattering object some particles will be scattered while others will pass on undisturbed, and we can therefore define a

probability that scattering will take place. Clearly the probability of scattering will be proportional to the flux density of the incident beam, and the probability that a particle will be scattered from a beam of unit flux density—that is, where one particle passes through unit area per second—in unit time is known as the scattering cross section, σ. It follows that σ has the dimensions of area and, in the classical case of a beam of small particles that interact with a scattering object only when they strike it, it is equal to the geometrical cross section of the body in a plane perpendicular to the beam. In quantum mechanics, however, such a simple interpretation is rarely possible.

We are often interested, not only in the total probability of scattering, but also in the probability that the particles are scattered in a particular direction, which is usually defined by the spherical polar angles θ and ϕ referred to the direction of the incident beam as the polar axis. To this end we define the *differential cross section* $\sigma(\theta, \phi)$ such that, for an incident beam of unit flux density, $\sigma(\theta, \phi)\, d\Omega$ is the probability per second of a particle being scattered into the element of solid angle $d\Omega$ around the direction defined by θ and ϕ. If we express $d\Omega$ in terms of θ and ϕ in the usual way

$$d\Omega = \sin\theta \, d\theta \, d\phi \tag{9.23}$$

It follows directly from the definitions of the total and differential cross sections that

$$\begin{aligned} \sigma &= \int \sigma(\theta, \phi)\, d\Omega \\ &= \iint \sigma(\theta, \phi) \sin\theta \, d\theta \, d\phi \end{aligned} \tag{9.24}$$

In many cases the problem has cylindrical symmetry about an axis parallel to the incident beam and the differential cross section is consequently independent of ϕ and written as $\sigma(\theta)$. In this case the probability per second of scattering into the element of angle $d\theta$ around θ, irrespective of the value of ϕ, is given by $2\pi\sigma(\theta) \sin\theta \, d\theta$.

Centre-of-mass frame

Finally in this section we note that scattering experiments are often carried out using beams of particles whose mass is comparable to the mass of the particles in the target (e.g. the scattering of neutrons by the protons in hydrogen referred to near the end of this chapter). As is shown in chapter 10, such a situation can be treated by considering the corresponding case of a particle of mass μ equal to the reduced mass of the two particles—that is, $\mu = m_1 m_2/(m_1 + m_2)$—interacting with a fixed target. We can therefore apply all the theory relating to the scattering of particles from fixed objects to this case and obtain results referred to a frame of reference attached to the centre of mass of the system, which may then be transferred back into the laboratory frame by standard methods.

9.3 The Born approximation

We saw in the earlier discussion of one-dimensional scattering that the energy eigenfunction of the system could be expressed as linear combinations of incident and scattered waves, and the scattering probability could then be calculated from the expansion coefficients. A similar procedure can, in principle, be followed in the three-dimensional case, but the initial solution of the time-independent Schrödinger equation for the energy eigenfunctions is generally difficult and often impossible. In the next section we shall discuss a method for doing this in the case of a spherically symmetric scattering potential, but for the moment we shall describe a procedure known as the *Born approximation*. This method is based on first-order perturbation theory and is usually valid when the average energy of the interaction between an incident particle and the scatterer is much smaller than the particle's kinetic energy.

We consider the case of a beam of particles approaching a scattering object along a direction parallel to the vector $\mathbf{k}_0$. Except in the vicinity of the scatterer, the potential is zero and the incident beam can therefore be described by a plane wavefunction whose time-independent part is u_0 where

$$u_0 = V^{-1/2} \exp(i\mathbf{k}_0 \cdot \mathbf{r}) \tag{9.25}$$

We have assumed this wavefunction to be normalized so that there is, on average, one particle in the volume V, and we impose periodic boundary conditions, similar to those applied in the one-dimensional case. The allowed values of the Cartesian components of $\mathbf{k}_0$ are therefore

$$k_{0x} = 2n_1\pi/L_1 \qquad k_{0y} = 2n_2\pi/L_2 \qquad k_{0z} = 2n_3\pi/L_3 \tag{9.26}$$

where n_1, n_2, and n_3 are integers and $L_1L_2L_3 = V$. We now imagine that scattering has taken place and that a scattered particle has been detected moving in some direction $\mathbf{k}$ (cf. figure 9.2). The normalized wavefunction will now be u_1 where

$$u_1 = V^{-1/2} \exp(i\mathbf{k} \cdot \mathbf{r}) \tag{9.27}$$

and boundary conditions similar to (9.26) are imposed on the Cartesian components of $\mathbf{k}$. Remembering that the scattering potential is to be treated as a perturbation, we see that u_0 and u_1 are both energy eigenfunctions of the unperturbed (zero-potential everywhere) problem. We can therefore represent the scattering process as a *transition* from the state u_0 to the state u_1 and relate the scattering probability and hence the scattering cross section to the transition probability.

We previously calculated an expression for the transition probability when we considered time-dependent perturbation theory in chapter 8. Although the potential is not time dependent in the present case, an identical argument can be applied, after putting $\omega = 0$, as far as equation (8.29). At this point we see that transitions occur only if the energies of the initial and final states are identical

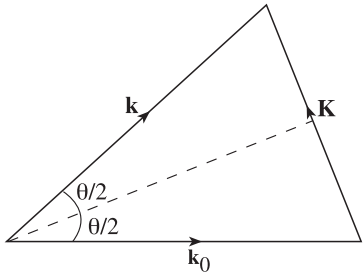


Figure 9.2. The relationship between the scattering angle (θ), the scattering vector (**K**), and the wavevectors of the incident ($\mathbf{k}_0$) and scattered (**k**) waves.

(that is, if $\omega_{mn} = 0$, so that $|\mathbf{k}_0| = |\mathbf{k}|$ and the scattering is elastic) when *both* of the terms on the right-hand side of (8.29) are simultaneously non-zero. This introduces a factor of two into the subsequent equations and a consequent factor of four into the Fermi-golden-rule expression (8.34) for the transition rate W which is therefore

$$W = \frac{2\pi}{\hbar^2}|U_{\mathbf{k}\mathbf{k}_0}|^2 g(\omega) \tag{9.28}$$

The matrix element $U_{\mathbf{k}\mathbf{k}_0}$ is calculated using the approximate eigenfunctions u_0 and u_1 and the scattering potential $U(\mathbf{r})$ (note that we now use U to represent the potential in order to avoid confusion with the volume V):

$$\begin{aligned} U_{\mathbf{k}\mathbf{k}_0} &= \int u_1^* U(\mathbf{r}) u_0 \, d\tau \\ &= \frac{1}{V}\int U(\mathbf{r}) \exp(-i\mathbf{K}\cdot\mathbf{r}) \, d\tau \end{aligned} \tag{9.29}$$

where $\mathbf{K} = \mathbf{k} - \mathbf{k}_0$ and is known as the *scattering vector*.

To calculate the differential cross section using (9.28) we must obtain an expression for the number of states in the angular frequency range ω to $\omega + d\omega$ available to a particle scattered into the element of solid angle $d\Omega$. It follows directly from the restrictions on the allowed values of **k** (9.26) that the number of states in an element of volume $d^3\mathbf{k} = dk_x \, dk_y \, dk_z$ in k-space is equal to

$$(L_1 L_2 L_3/8\pi^3)\, d^3\mathbf{k} = (V/8\pi^3)\, d^3\mathbf{k} \tag{9.30}$$

where we have assumed that V is large enough for the volume of the element $d^3\mathbf{k}$ to be large compared with that of the k-space cell, $8\pi^3/V$. If we now consider the states whose wavevectors lie in the element of solid angle $d\Omega$ around the direction **k** and whose magnitudes lie between k and $k + dk$, these occupy a k-space volume equal to $k^2 \, d\Omega \, dk$ and the total number of such states is therefore

$$(V/8\pi^3)k^2 \, d\Omega \, dk \tag{9.31}$$

But we know that the energy of a free particle is $E = \hbar^2 k^2/2m$ so that

$$dE = (\hbar^2 k/m)\, dk$$

and therefore $d\omega = (\hbar k/m)\, dk$ where $E = \hbar\omega$. Writing the required density of states as dg to emphasize the fact that it refers to an element of solid angle $d\Omega$, the number of states with angular frequency between ω and $\omega + d\omega$, and whose k-vectors lie within $d\Omega$, equals $dg\, d\omega$ and we have

$$dg = \frac{mkV}{8\pi^3\hbar}\, d\Omega \tag{9.32}$$

We can now obtain an expression for the probability per unit time (dW) of a particle being scattered from a state u_0 to a state whose wavevector lies within the element of solid angle $d\Omega$ around the direction $\mathbf{k}$, by substituting the expression (9.32) for dg in place of g in (9.28) and using (9.29):

$$dW = \frac{mk}{4\pi^2 V\hbar^3}\left|\int U(\mathbf{r})e^{-i\mathbf{K}\cdot\mathbf{r}}\, d\tau\right|^2 d\Omega \tag{9.33}$$

Finally, to calculate the differential cross section we must divide the right-hand side of (9.33) by the magnitude of the incident flux ($\hbar k/mV$) and by $d\Omega$ to give

$$\sigma(\theta, \phi) = \frac{m^2}{4\pi^2\hbar^4}\left|\int U(\mathbf{r})e^{-i\mathbf{K}\cdot\mathbf{r}}\, d\tau\right|^2 \tag{9.34}$$

Thus, provided we know the form of the scattering potential, the scattering cross section can be calculated by evaluating the Fourier transform of $U(\mathbf{r})$ and hence the right-hand side of (9.34). This result is equivalent to that which would be obtained on the basis of Fraunhofer diffraction theory used in optics, where the amplitude of the scattered light as a function of scattering vector is proportional to the Fourier transform of the diffracting object.

We shall shortly discuss a couple of examples of the application of the Born approximation when the scattering potential is spherically symmetric, in which case the volume integral in (9.34) can be partly evaluated using spherical polar coordinates r, θ' and ϕ' referred to the direction of $\mathbf{K}$ as polar axis. (NB: θ' and ϕ' should be distinguished from θ and ϕ which define the direction of $\mathbf{k}$ relative to $\mathbf{k}_0$.) Thus

$$\begin{aligned}\int U(r)e^{-i\mathbf{K}\cdot\mathbf{r}}\, d\tau &= \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} U(r)e^{-iKr\cos\theta'} r^2\, dr\, \sin\theta'\, d\theta'\, d\phi' \\ &= \frac{4\pi}{K}\int_0^{\infty} U(r) r \sin Kr\, dr\end{aligned} \tag{9.35}$$

and the differential cross section becomes

$$\sigma(\theta) = \frac{4m^2}{\hbar^4 K^2}\left|\int_0^{\infty} U(r) r \sin Kr\, dr\right|^2 \tag{9.36}$$

We note that the differential cross section is independent of ϕ and that its dependence on θ is through the magnitude of the scattering vector K which is seen from figure 9.2 to be equal to $2k_0 \sin(\theta/2)$.

Example 9.1 Scattering by a spherical potential well or step Our first example relates to scattering by a spherically symmetric potential given by

$$\left.\begin{aligned} U(r) &= 0 \qquad r > a \\ &= U_0 \qquad r \leqslant a \end{aligned}\right\} \tag{9.37}$$

where U_0 and a are constants. U_0 may be negative (potential well) or positive (potential step) but, in any case, it is to be assumed small enough for the Born approximation to be applied. The more general case including larger U_0 will be treated by the method of partial waves towards the end of this chapter. The potential is spherically symmetric so (9.36) can be used and the integral on the right-hand side of this equation is

$$U_0 \int_0^a r \sin Kr \, dr = U_0(\sin Ka - Ka \cos Ka)/K^2 \tag{9.38}$$

This can be substituted into (9.36) to produce an expression for the scattering cross section.

In the particular case where Ka is much less than one, the trigonometric terms can be expanded as series of ascending powers of Ka and the first non-vanishing term in (9.38) is equal to $\frac{1}{3}U_0 K a^3$. The differential cross section in this limit is then

$$\sigma(\theta) = \frac{4m^2 U_0^2 a^6}{9\hbar^4} \tag{9.39}$$

Thus, if the geometrical radius of the scatterer is much less than the wavelength ($\lambda = 2\pi/k$) associated with the incident particles, Ka will be small for all scattering angles, the scattering will be isotropic (i.e. independent of θ), and the total cross section will be simply 4π times that given in (9.39). The fact that objects whose dimensions are much less than λ scatter isotropically also follows more generally from equation (9.34) where we see that if $\mathbf{K} \cdot \mathbf{r}$ is much less than one for all values of $\mathbf{r}$ for which U is neither zero nor insignificantly small, then the differential cross section is independent of angle. It is also in agreement with the familiar phenomenon in optics in which light striking an object whose dimensions are much smaller than the wavelength is scattered equally in all directions.

Example 9.2 Rutherford scattering We now consider scattering by a potential whose magnitude is inversely proportional to the distance from the origins of the particle being scattered. That is,

$$U(r) = \beta r^{-1} \tag{9.40}$$

where β is a constant. This is known as Rutherford scattering because it can be applied to the case of positively charged α particles approaching a target containing atoms with positively charged nuclei (when $\beta = 2Ze^2/4\pi\varepsilon_0$), and this arrangement was first studied by Rutherford in 1911.

Substituting from (9.40) the integral in (9.36) becomes

$$\beta \int_0^\infty \sin Kr \, dr = -\frac{\beta}{K}[\cos Kr]_0^\infty \tag{9.41}$$

The integral is indeterminate at its upper limit, but in practice the Coulomb potential is always modified at large distances (e.g. due to screening by the atomic electrons) so that at large r it falls off more rapidly than is implied by (9.40); the integral is then negligibly small at large r and we obtain the following expression for the differential cross section:

$$\begin{aligned} \sigma(\theta) &= \frac{4m^2\beta^2}{\hbar^4 K^4} \\ &= \frac{m^2\beta^2}{4\hbar^4 k_0^4 \sin^4(\theta/2)} \\ &= \frac{m^2\beta^2}{4p_0^4 \sin^4(\theta/2)} \end{aligned} \tag{9.42}$$

where p_0 is the magnitude of the momentum of the incident particles and θ is the angle between the incident and scattered directions as before. It was Rutherford's confirmation of this formula by the observation of the scattering of α particles from gold foil that led to his postulate of the nuclear atom; of course this development preceded quantum mechanics, but by a happy coincidence it turns out that the classical and quantum expressions are identical in this case.

9.4 Partial wave analysis

In the final section of this chapter we consider an approach to the scattering problem which is, in some ways, the opposite to that adopted in the Born approximation. Whereas the latter expresses the cross section in terms of the Fourier transform of the scattering potential—that is, the potential is expanded as a linear combination of plane waves whose form matches that of the wavefunctions of the incident and scattered particles—in partial wave analysis we start with the eigenfunctions of the scattering potential and express the incident plane wave as a linear combination of these. This method is particularly useful if the scattering potential is spherically symmetric, and this is the only case we shall consider. Unlike the Born approximation, no limitations are placed on the strength of the interaction between the particle beam and the scatterer.

The principle of the method is similar to that adopted for the one-dimensional example discussed earlier. We first obtain the wavefunction as an energy

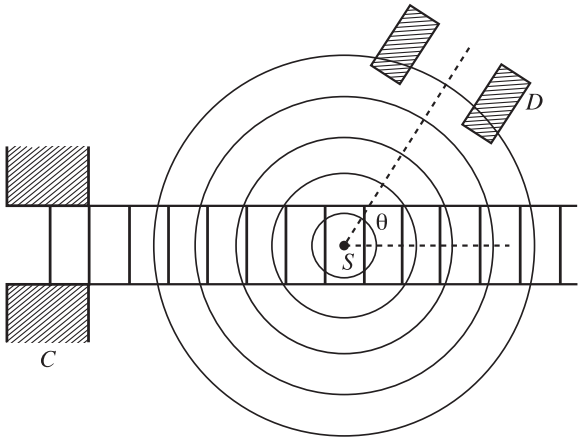


Figure 9.3. Particles passing through the collimator C are represented by a plane wave which is scattered by the scatterer S. Only the scattered wave enters the detector D, but there is a region around the scatterer, much larger than the scatterer itself, over which the wavefunction is a linear combination of the incident and scattered waves.

eigenfunction whose eigenvalue equals the energy of the incident particles. This is then expressed as a sum of a plane wave representing the incident beam and a scattered wave, and we use the measurement postulate to calculate the differential cross section as the ratio of the flux of the scattered wave through an element of solid angle in a particular direction, to the flux density of the incident wave. For this procedure to be applicable, the experimental measurement must distinguish between the incident and scattered waves and a typical set-up to achieve this is shown in figure 9.3. The incident beam is defined by a collimator whose diameter is small compared with the distance from the scatterer to the detector, so ensuring that particles cannot enter the detector without first being scattered, and that the wavefunction in the region of the detector is therefore just that of the scattered wave. However, the collimator diameter must be much larger than both the wavelength of the incident beam and the dimensions of the scattering object. This ensures both that the incident particles can be represented by a plane wave, and also that there is a region around the scatterer, much larger than the scatterer itself, over which the wavefunction is a linear combination of the incident wave and scattered waves.

We begin our discussion by considering the form of the energy eigenfunctions associated with a spherically symmetric potential $U(r)$. This problem was discussed in chapter 3, but at that point we restricted our attention to bound states of potential wells where the total energy was found to be quantized. In the scattering case, however, we are interested in wavefunctions representing particles approaching the scatterer from a large distance and leaving it again.

Thus, either the potential is such that binding is impossible (e.g. if it is repulsive at all distances) or, if bound states do exist, the incident energy is too great for binding to occur. This means that the energy levels of the scattering system are not confined to a discrete set of values.

To obtain the energy eigenfunctions of eigenvalue E for a spherically symmetric potential $U(r)$, we must solve the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m}\nabla^2 u + U(r)u = Eu \tag{9.43}$$

This equation can be separated into spherical polar coordinates just as was done for a bound particle in chapter 3, by putting

$$u = R(r)Y_{lm}(\theta, \phi) \tag{9.44}$$

where the Y_{lm} are spherical harmonics which are defined in that chapter and R is obtained by solving the radial equation

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{l(l+1)\hbar^2}{2mr^2}R + UR = ER \tag{9.45}$$

If we now define k and $\chi(r)$ so that $R = \chi/r$ and $E = \hbar^2k^2/2m$, (9.45) becomes

$$\frac{d^2\chi}{dr^2} + \left(k^2 - \frac{l(l+1)}{r^2} - \frac{2mU}{\hbar^2}\right)\chi = 0 \tag{9.46}$$

In a scattering experiment we are interested in the wavefunction at a large distance from the scattering object. We can therefore obtain a lot of information from the asymptotic form of the solutions at large r when we can neglect the last two terms in the bracket—assuming that the potential $U(r)$ is effectively zero at large r, which is nearly always the case. Equation (9.46) then becomes

$$\frac{d^2\chi}{dr^2} + k^2\chi = 0 \tag{9.47}$$

the general solution to which can be written as

$$\chi = Ae^{ikr} + Be^{-ikr} \tag{9.48}$$

where A and B are constants. Using (9.44) we obtain the following expressions for the asymptotic form of the energy eigenfunctions which we now write as u_{klm}:

$$u_{klm} = \frac{1}{r}(Ae^{ikr} + Be^{-ikr})Y_{lm}(\theta, \phi) \tag{9.49}$$

The energy eigenvalues are independent of the quantum numbers l and m, so any linear combination of the functions u_{klm} that have the same value of k is also an eigenfunction. We shall shortly obtain such a linear combination

with coefficients chosen so that it has the desired form of a sum of the incident plane wave and a scattered wave, but in the meantime useful information can be obtained by considering the case where the wavefunction is just one of the eigenfunctions (9.49). The first term in this expression represents a wave travelling radially outwards while the second corresponds to a wave travelling inwards. This wavefunction would therefore apply to an experiment where particles approached the scatterer along the radial direction with the numbers coming in at different angles (θ, ϕ) proportional to $|Y_{lm}|^2$. The spherical harmonics, Y_{lm} are illustrated in figure 3.3 and we see, for example, that if $l = m = 0$ all directions of approach are equally probable, while if $l = 1$ and $m = 0$, the most favoured direction of approach is along the z axis. For such an experiment, equation (9.49) tells us that the angular part of the wavefunction after the scattering will be identical to that beforehand. The total number of particles (N_1) passing outwards per second through the surface of a sphere of radius r centred on the origin is readily obtained by substituting the first term in (9.49) (which we refer to below as u_1) into the general expression for the flux (9.20) and integrating over all solid angles, remembering that the radial component of $\nabla\psi$ is $\partial\psi/\partial r$. Thus

$$\begin{aligned} N_1 &= -\frac{i\hbar}{2m}\int_0^{2\pi}\int_0^{\pi}\left(u_1^*\frac{\partial u_1}{\partial r} - u_1\frac{\partial u_1^*}{\partial r}\right)r^2\sin\theta\, d\theta\, d\phi \\ &= -\frac{i\hbar}{2m}\left(2|A|^2\frac{ik}{r^2}\right)r^2\int_0^{2\pi}\int_0^{\pi}|Y_{lm}|^2\sin\theta\, d\theta\, d\phi \\ &= \frac{\hbar k}{m}|A|^2 \end{aligned} \tag{9.50}$$

using the fact that the spherical harmonics are normalized. We note that N_1 is independent of r as it must be because no particles are lost or gained as the distance from the scatterer is varied. The number flowing inward per second, N_2, is similarly calculated using the second term in (9.49) as

$$N_2 = \frac{\hbar k}{m}|B|^2 \tag{9.51}$$

We now note that in a spherically symmetric system the energy eigenfunctions (9.49) are also eigenfunctions of the total angular momentum and its z component so that these two quantities must be conserved in an energy eigenstate. Thus no change in the quantum numbers l and m can occur as a result of the scattering process and all particles incident on the scatterer in a state described by the second term in (9.49) must leave in a state described by the first term. It follows that N_1 and N_2 must be equal so that the constants A and B can differ only by a phase factor. We can therefore write

$$u_{klm} = \frac{A}{r}[e^{-ikr} - e^{i(kr-l\pi+2\delta_{lm})}]Y_{lm}(\theta,\phi) \tag{9.52}$$

where we have defined the *phase shift* δ_{lm} (which is *not* the Kronecker delta in this case) so that the phase factor relating B to A is $-\exp i(2\delta_{lm} - l\pi)$ as this form is particularly convenient for the later discussion. We conclude that, if an experiment were performed in which the incident particles had a wavefunction of the form

$$\frac{A}{r}e^{-ikr}Y_{lm}(\theta, \phi) \tag{9.53}$$

then all the particles would be scattered and the angular part of the scattered wave would be identical to that of the incident wave apart from a phase factor. The actual value of the phase shift in any particular case is determined by the form of the scattering potential. We note the similarity between this result and that obtained in our earlier example of scattering by a potential step in one dimension (cf. (9.15)).

In practice, of course, scattering experiments are not performed with incident waves of the form (9.53) which represent particles approaching the scatterer from all sides with an angular distribution determined by the spherical harmonic. Instead, as we have seen earlier, beams of particles are used all of which are travelling in the same (say the z) direction so that the incident wavefunction u_{k0} has the plane-wave form

$$\begin{aligned} u_{k0} &= V^{-1/2}\exp(ikz) \\ &= V^{-1/2}\exp(ikr\cos\theta) \end{aligned} \tag{9.54}$$

in spherical polar coordinates. In order to proceed further we have to express the plane wave (9.54) as a linear combination of incoming and outgoing spherical waves of the form discussed earlier. This is a standard mathematical expression (see, for example, G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1958) which in the asymptotic limit of large r, has the form

$$V^{-1/2}e^{ikr\cos\theta} = \frac{1}{2}V^{-1/2}\sum_{l=0}^{\infty}(2l+1)i^{2l+1}\left[\frac{1}{kr}e^{-ikr} - \frac{1}{kr}e^{i(kr-l\pi)}\right]P_l(\cos\theta) \tag{9.55}$$

where $P_l(\cos\theta)$ is the Legendre polynomial of order l and we remember from chapter 3 that this is related to the spherical harmonic Y_{l0} by

$$Y_{l0} = \left(\frac{2l+1}{4\pi}\right)^{1/2}P_l(\cos\theta) \tag{9.56}$$

If we compare (9.55) and (9.56) with (9.52) we see that a plane wave travelling in the positive z direction can be expressed as a linear combination of incoming and outgoing spherical waves—known as 'partial waves'—which all have $m = 0$ and different values of l: the phase relation between the incoming and outgoing partial waves in this case is such that the phase shifts δ_{lm} in (9.52) are all equal to zero.

Given that (9.55) is a representation of the incident plane wave we can now consider how this is modified by the presence of the scattering object. We remember that, in the steady state, the total wavefunction must be an energy eigenfunction and must therefore be equal to a linear combination of the eigenfunctions determined earlier (9.52) that have the same value of k as the incident wave. Thus

$$\psi = \sum_{lm} \frac{A_{lm}}{r}[e^{-ikr} - e^{i(kr-l\pi+2\delta_{lm})}]Y_{lm}(\theta, \phi) \tag{9.57}$$

However, the total wavefunction can also be written as a sum of an incident and a scattered wave and we know that when these are separated in the course of the experiment (see figure 9.3) the latter represents only particles which are moving outwards from the scatterer. The scattered wave therefore must not contain any components with negative values of the exponent ikr and the coefficients of such terms in the total wavefunction (9.57) must be the same as in the incident wave (9.55). Thus we have, using (9.56),

$$A_{lm} = 0 \qquad m \neq 0$$
$$A_{l0} = \tfrac{1}{2}[4\pi(2l+1)/V]^{1/2} i^{2l+1} k^{-1}$$

The first of these conditions confirms that the wavefunction is independent of ϕ, as would be expected from the symmetry of the problem. If we now substitute back into (9.57) and make use of (9.56) again we get

$$\psi = \frac{1}{2}V^{-1/2}\sum_{l=0}^{\infty}(2l+1)i^{2l+1}\left[\frac{1}{kr}e^{-ikr} - \frac{1}{kr}e^{i(kr+2\delta_l-l\pi)}\right]P_l(\cos\theta) \tag{9.58}$$

where we have rewritten δ_{lm} as δ_l, because (9.58) contains only terms with $m = 0$. We now subtract the incident wave (9.55) from the total wavefunction (9.58) to get an expression for the scattered wavefunction ψ_s which then has the form

$$\psi_s = V^{-1/2}\frac{1}{r}e^{ikr}f(\theta) \tag{9.59}$$

where

$$\begin{aligned} f(\theta) &= \frac{1}{2k}\sum_{l=0}^{\infty}(2l+1)i^{2l+1}e^{-il\pi}(1 - e^{2i\delta_l})P_l(\cos\theta) \\ &= \frac{1}{k}\sum_{l=0}^{\infty}(2l+1)e^{i\delta_l}\sin\delta_l P_l(\cos\theta) \end{aligned} \tag{9.60}$$

We note that if all the phase shifts δ_l are zero, $f(\theta)$ vanishes and there is no scattering, which is consistent with the fact that the total wavefunction (9.58) is

then identical to the incident plane wave (9.55). The extent and nature of the scattering can therefore be calculated, using the plane shifts δ_l, and we shall shortly discuss how to evaluate these in particular cases. First, however, we use the results derived so far to obtain expressions for the differential and total scattering cross sections.

To evaluate the differential cross section we must calculate the number (dN) of scattered particles crossing an area $r^2\,d\Omega$ on the surface of a sphere of radius r per second. Using the general expression for the flux (9.19) and remembering that the radial component of $\nabla\psi_s$ is $\partial\psi_s/\partial r$ we get

$$dN = \frac{\hbar k}{mV}|f(\theta)|^2\,d\Omega \tag{9.61}$$

The differential cross section $\sigma(\theta)$ is obtained by dividing dN by $d\Omega$ and by the incident flux $\hbar k/mV$ (cf. (9.22)). Thus

$$\begin{aligned}\sigma(\theta) &= |f(\theta)|^2 \\ &= \frac{1}{k^2}\left|\sum_{l=0}^{\infty}(2l+1)e^{i\delta_l}\sin\delta_l P_l(\cos\theta)\right|^2\end{aligned} \tag{9.62}$$

The total cross section σ is obtained by integrating (9.62) over all solid angles

$$\begin{aligned}\sigma &= 2\pi\int_0^{\pi}\sigma(\theta)\sin\theta\,d\theta \\ &= \frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\sin^2\delta_l\end{aligned} \tag{9.63}$$

where we have used the expression

$$\int_0^{\pi} P_l P_{l'}\sin\theta\,d\theta = \frac{2l}{2l+l}\delta_{ll'} \tag{9.64}$$

($\delta_{ll'}$ now being the Kronecker delta) which follows from (9.56) and the fact that the spherical harmonics are orthonormal.

Thus, if the phase shifts δ_l are known, the differential and total cross sections can be calculated using (9.62) and (9.63). The phase shifts are determined by the detailed size and shape of the scattering potential and their calculation is often a long and complicated process which is only practicable using numerical methods. However, in some cases all but a few of the phase shifts are zero or negligibly small and expressions for the scattering cross sections are comparatively simple. This is particularly true in the case of the scattering of low-energy particles when often the only significant contributions to (9.62) and (9.63) are from terms with $l = 0$. We can see why this is so from the following semi-classical argument. The partial waves correspond to particular values of the angular momentum L given by $\sqrt{l(l+1)}\hbar$; but classically a particle with this angular momentum and with

linear momentum $\hbar k$ must pass the origin at a distance x such that $L = \hbar kx$ and therefore

$$kx = \sqrt{l(l+1)}$$

We can conclude, therefore, that if the range of the scattering potential is of order a (that is, $U(r)$ is zero or negligibly small if $r > a$) the scattering of the lth partial wave will not contribute unless

$$ka \geqslant \sqrt{l(l+1)} \tag{9.65}$$

For sufficiently small k (i.e. for sufficiently low energies) (9.65) is satisfied only if $l = 0$ and this is the only partial wave to be significantly scattered. This case is often referred to as s-wave scattering (cf. the terminology for $l = 0$ bound states set out in chapter 3). It corresponds to a spherically symmetric wave and therefore isotropic scattering, in agreement with the result obtained earlier using the Born approximation in the special case where $ka \ll 1$. A rigorous quantum-mechanical proof of this involves investigating the form of the partial waves contributing to the incident plane wave near the origin, rather than in the asymptotic limit of large r. We shall not do this here, but simply quote the result that these are proportional to $(kr)^l$ in the limit $r \to 0$; in the case of small k the incident partial waves are therefore small over the volume of the scattering potential and no appreciable scattering results unless the condition (9.65) is satisfied.

We shall now discuss two examples of the application of the method of partial waves and in each case we shall confine our consideration to s-wave scattering.

Example 9.3 Scattering by hard spheres In this example we consider a scatterer whose radius is a and which cannot be penetrated by the incident particles. Thus the potential energy is given by

$$\left.\begin{aligned} U(r) &= \infty \qquad r \leqslant a \\ U(r) &= 0 \qquad r > a \end{aligned}\right\} \tag{9.66}$$

It follows that the wavefunction must be zero for $r \leqslant a$ and we remember that it is continuous at the boundary $r = a$.

If we consider s-wave scattering only, it follows from (9.46) that the asymptotic form of the wavefunction (9.52) is actually a solution to the Schrödinger equation for all values of r greater than a so that the continuity of the wavefunction at $r = a$ requires that

$$e^{-ika} - e^{i(ka+2\delta_0)} = 0$$

and the phase shift δ_0 is therefore given by

$$\delta_0 = -ka \tag{9.67}$$

The total cross section is obtained by substituting (9.67) into (9.63) to give

$$\sigma = \frac{4\pi \sin^2 ka}{k^2}$$

As the s-wave approximation is valid only in the limit $ka \ll 1$, we can write this as

$$\sigma \simeq 4\pi a^2 \tag{9.68}$$

Classically we would expect the scattering cross section for such a hard-sphere potential to be πa^2, so we have another example of how quantum mechanics produces results very different from those we would intuitively expect. In the next example, we shall see how a finite scattering potential can produce results that are even more dramatically different from classical expectations.

Example 9.4 Scattering by a potential well or step We now return to the problem previously treated using the Born approximation where the potential is given by

$$\left.\begin{aligned} U &= 0 \qquad r > a \\ U &= U_0 \qquad r \leqslant a \end{aligned}\right\} \tag{9.69}$$

As before, U_0 may be positive or negative, but now need not be small. We shall, however, consider explicitly only the case where the incident energy E is greater than U_0 and also confine our attention to s-wave scattering where $l = 0$ and $ka \ll 1$; it follows that our results will not be applicable to the repulsive case unless $U_0 \ll \hbar^2/2ma^2$.

In the region $r \leqslant a$ the general solution to the Schrödinger equation when $l = 0$ is readily seen to be

$$\frac{1}{r}(A' \sin k'r + B' \cos k'r) \tag{9.70}$$

where

$$k' = [2m(E - U_0)/\hbar^2]^{1/2} \tag{9.71}$$

and A' and B' are constants. However, the wavefunction must be finite everywhere, including the point $r = 0$, so the cosine term cannot exist and B' must be equal to zero. The $l = 0$ eigenfunction in the region $r > a$ again has the general form (9.52) and, if we apply the condition that both the wavefunction and its first spatial derivative must be continuous at $r = a$, we get

$$A[e^{-ika} - e^{i(ka+2\delta_0)}] = A' \sin k'a$$
$$ikA[-e^{-ika} - e^{i(ka+2\delta_0)}] = k'A' \cos k'a$$

If we divide the second of these equations by the first we get

$$\frac{ik[-e^{-ika} - e^{i(ka+2\delta_0)}]}{e^{-ika} - e^{i(ka+2\delta_0)}} = \frac{k' \cos k'a}{\sin k'a}$$

We can now multiply the numerator and denominator on the left-hand side by $\exp(-i\delta_0)$ and express the resulting expressions in trigonometric form as

$$k\cot(ka+\delta_0) = k'\cot k'a$$

which leads, after a little manipulation, to

$$\cot\delta_0 = \frac{k\tan ka\tan k'a + k'}{k\tan k'a - k'\tan ka} \tag{9.72}$$

$$\simeq \frac{k'\cot k'a}{k(1-k'a\cot k'a)} \tag{9.73}$$

where (9.73) is the limiting form of (9.72) when $ka \ll 1$. The scattering cross section for s-wave scattering is given by the $l=0$ term in (9.63) as

$$\sigma = \frac{4\pi}{k^2}\sin^2\delta_0$$

$$= \frac{4\pi}{k^2(1+\cot^2\delta_0)} \tag{9.74}$$

using standard trigonometric identities. Thus, given k, a, and U_0 we can calculate $\cot\delta_0$ and hence σ.

We note from (9.73) that if the magnitude of $k'a$ happens to be such that $k'a\cot k'a = 1$ (i.e. if $k'a = 4.49, 7.73, 10.9, \ldots$) $\cot\delta_0$ will be infinite, the cross section will be zero, and the incident particles will 'diffract past' the potential without being significantly scattered. Clearly this condition can be satisfied simultaneously with ka being small only if U_0 is negative—that is, if we are considering a potential well rather than a step. In such a case, the absence of scattering means that the wavefunction inside the well fits smoothly onto the plane wave outside. Clearly the potential well need not be square for this to occur, and the phenomenon has been observed experimentally in the case of the scattering of electrons by rare-gas atoms when it is known as the Ramsauer–Townsend effect: electrons of energy about 0.7 eV pass through helium gas without being significantly scattered.

The opposite extreme to this case occurs when the parameters are such that $k'a$ is approximately equal to an odd multiple of $\pi/2$ so that $\cot k'a$ is small. We can then ignore the second term in the denominator of (9.73) and write the cross section as

$$\sigma = \frac{4\pi}{k^2 + k'^2\cot^2 k'a} \tag{9.75}$$

and we note that, as ka and $\cot k'a$ are both much less than one, the cross section (9.75) is much greater than the geometrical cross section πa^2. This case is known as *resonant scattering*. It is interesting to compare the condition for such an

s-wave resonance with the equation governing the energy levels of a particle in a spherical potential well (cf. problem 3.5) which is

$$-k'' \cot k'' a = (-2mE_b/\hbar^2)^{1/2} \tag{9.76}$$

where

$$k'' = [2m(-U_0 - E)/\hbar^2]^{1/2}$$

and E_b is the binding energy of the system. In the case where E_b is small, k'' is nearly equal to k' and (9.76) is approximately equivalent to the resonance condition. We can conclude, therefore, that s-wave resonant scattering with large cross section is to be expected if there is a bound state of the system particle-plus-scatterer whose energy level lies just below the top of the well; the cross section (9.75) can then be expressed as

$$\sigma \simeq \frac{2\pi\hbar^2}{m(E - E_b)} \tag{9.77}$$

We note that as the energy E of the incident particles is decreased, the cross section approaches a limiting value determined by the binding energy. Clearly these conditions will also hold and resonance scattering will also occur if U_0 is just a little smaller than would be required for there to be a zero-energy bound state. In this case the quantity E_b measured from such scattering experiments is interpreted as the energy associated with a 'virtual' energy level. As with the Ramsauer–Townsend effect, we expect these results to be quite general and independent of the detailed shape of the well. The large phase shift can be explained semi-classically if we imagine the incident particle to be 'trapped' in the well for a time before it is re-emitted, and the closer the energy of the particle is to that of the bound state, the greater will be the probability of trapping. An example of resonant scattering is that of low-energy neutrons from hydrogen where the s-wave cross section of 20.4×10^{-28} m^2 is more than one hundred times greater than would be expected classically, given the range of the neutron–proton interaction (2×10^{-15} m); and indeed the deuteron (which consists of a proton and neutron bound together) does have an energy level with a very small binding energy.

Problems

9.1 Particles of energy E move in one dimension towards a potential barrier of height U_n and width a. Show that if E is greater than U_0 the probability of scattering P is given by

$$P = \left[1 + \frac{4E(E - U_0)}{U_0^2 \sin^2 k_2 a}\right]^{-1}$$

where $k_2 = [2m(E - U_0)/\hbar^2]^{1/2}$, and obtain a corresponding expression for the transmission probability.

9.2 Show that the scattering probability derived in problem 9.1 is zero if $k_2 a = n\pi$ where n is an integer and has a maximum value if the incident energy is such that $k_2 a \cot k_2 a = 2 - U_0/E$.

Compare this situation and that set out in problem 9.1 with those applying in the Ramsauer–Townsend effect and in s-wave resonant scattering.

9.3 Obtain expressions for the scattering and transmission probabilities for the system described in problem 9.1, but now assume E to be less than U_0. Use your answers to all three questions to draw graphs of the scattering probability as a function of E for values of E between 0 and $5U_0$ in the cases where $U_0a^2 = \pm 10\hbar^2/m$.

9.4 Show that in one dimension the scattering probability calculated using the Born approximation is P' where

$$P' = \frac{m^2}{\hbar^4 k^2}\left|\int_{-\infty}^{\infty} U(x)e^{-2ikx}\,dx\right|^2$$

Use this expression to calculate the scattering probability for the system described in problem 9.1 and show that it is equivalent to the expression given there in the limit $E \gg U_0$.

9.5 Particles of momentum $\hbar k_0$ travel along the z axis towards a three-dimensional rectangular potential well of depth U_0 and dimensions $2a \times 2b \times 2c$ where a, b and c are parallel to the x, y and z axes respectively. Use the Born approximation to show that, if the incident energy is much greater than U_0, the differential scattering cross section in the direction (k_x, k_y, k_z) is equal to

$$\frac{16m^2U_0^2\sin^2(k_xa)\sin^2(k_yb)\sin^2[(k_z-k_0)c]}{\pi^2\hbar^4k_x^2k_y^2(k_z-k_0)^2}$$

9.6 Use the Born approximation to estimate the differential scattering cross section when the scattering potential is spherically symmetric and has the form Ar^{-2} where A is a constant.

Hint: $\int_0^\infty x^{-1}\sin x\,dx = \pi/2$.

9.7 Show that the expression for s-wave scattering by a spherical potential well—see equations (9.72) and (9.74)—goes over to that obtained from the Born approximation (cf. (9.39)) in the limit where $E \gg U_0$ and $ka \ll 1$.

Hint: do not use (9.73), but obtain a limit of (9.72) when k and k' are both small.

9.8 Show that the phase shift of the $l = 0$ partial wave in the case of scattering from a potential step, whose height U_0 is greater than the energy E of the incident particles, is given by equation (9.72) with $\tan k'a$ replaced by $\tanh k'a$ wherever it appears. Show that in the limit $U_0 \to \infty$ this result is consistent with that obtained earlier in the case of scattering by hard spheres.

9.9 Show that the s-wave cross section for very low-energy neutrons scattered by protons is about 2.4×10^{-28} m^2, on the assumption that the binding energy of the deuteron is -2.23 MeV (remember to use reduced mass). This is considerably smaller than the observed cross section of 20.4×10^{-28} m^2 because of the spin dependence of the neutron–proton interaction. In the ground state of the deuteron the z components of the two spins are parallel, but this is true for only three-quarters of all scattering events. Show that the experimental cross section can be reproduced if the state with zero total spin has a virtual level with energy about 70 keV.

Chapter 10

Many-particle systems

Although many of the principles of quantum mechanics can be adequately illustrated by considering systems that consist of only one particle subject to external forces, there are a number of important phenomena that are manifest only in systems containing two or more particles, and these will be the subject of the present chapter. We shall begin with some general statements and go on to consider the case of two interacting particles subject to no external forces, when we shall find that the problem can be separated into one describing the behaviour of the centre of mass of the system and another describing the relative motion. We shall then consider the case of two non-interacting particles and show that the particles can often be treated independently as would be expected. In the case of indistinguishable particles, however, we shall find that a symmetry is imposed on the wavefunction, which ensures that the behaviour of such particles is coupled even when they do not interact. We shall introduce the problem of interacting systems containing more than two identical particles, explaining what is meant by 'fermions' and 'bosons' and how their properties compare. These results will be illustrated by discussing the energy levels and optical spectra of the helium atom, starting with the approximation that the two electrons do not interact and extending the discussion to the realistic case using perturbation theory. The chapter closes with a consideration of the effects of particle indistinguishability on scattering theory.

10.1 General considerations

A system containing N particles is described by a wavefunction that is, in general, a function of the positions and spins of all the particles and of time. Ignoring spin for the present, we write this wavefunction as $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, t)$ and we can straightforwardly extend the probabilistic interpretation set up in the earlier chapters for a single particle, so that

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, t)|^2 \, d\tau_1 \, d\tau_2 \ldots d\tau_N$$

is the probability that particle 1 will be found in the element of volume $d\tau_1$ in the vicinity of $\mathbf{r}_1$, simultaneously with particle 2 being found within $d\tau_2$ in the vicinity of $\mathbf{r}_2$, etc. Probability distributions for other dynamical quantities can be derived from Ψ by the general procedure described in chapter 4, given the form of the appropriate operators and assuming that the corresponding eigenvalue equations can be solved. In general, integrals occurring in the basic theoretical expressions are with respect to all the coordinates of all the particles in the system. For example, the normalization condition becomes

$$\int \dots \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, t)|^2 \, d\tau_1 \, d\tau_2 \dots d\tau_N = 1 \tag{10.1}$$

The operators representing measurable quantities may be specific to particular particles or may represent global properties of the system. Thus the operator representing the x component of the momentum of the ith particle is $\hat{P}_{xi} = -i\hbar\partial/\partial x_i$, while that representing the x component of the total momentum of a system of N particles is

$$\hat{P}_x = \sum_{i=1,N} \hat{P}_{xi} = -i\hbar \sum_{i=1,N} \partial/\partial x_i \tag{10.2}$$

As usual, we shall be particularly interested in the energy eigenvalues and eigenfunctions, and the Hamiltonian operator representing the total energy of a system of N particles has the following general form (omitting possible spin-dependent terms)

$$\hat{H} = -\sum_{i=1,N} \frac{\hbar^2}{2m_i} \nabla_i^2 + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \tag{10.3}$$

where m_i is the mass of the ith particle, ∇_i is the vector operator differentiating with respect to the coordinates of the ith particle, and V is the potential energy of the system which, in general, will include contributions from external forces along with the energy associated with the interactions between the particles. The eigenvalue equation corresponding to the Hamiltonian (10.3) is therefore a partial differential equation containing $3N$ variables. We saw in the early chapters how difficult the solution of a system containing only the variables associated with a single particle could be, so it is not surprising that exact solutions to the many-body problem are possible only in simple cases. In much of the ensuing discussion we shall refer to systems containing only two particles, because these turn out to illustrate many of the important features of many-body systems, but we shall also refer to the properties of 'many-body' systems containing a large number of particles.

10.2 Isolated systems

In this section we consider a system of two particles that are not subject to any external forces so that the potential energy depends only on the relative position

of the particles and can be written as $V(\mathbf{r}_1 - \mathbf{r}_2)$. The Hamiltonian is then

$$\hat{H} = -\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + V(\mathbf{r}_1 - \mathbf{r}_2) \tag{10.4}$$

We now change variables from the particle positions $\mathbf{r}_1$ and $\mathbf{r}_2$ to those of the centre of mass of the system, $\mathbf{R}$ and their relative position, $\mathbf{r}$. That is

$$\left.\begin{aligned} \mathbf{R} &= \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \\ \mathbf{r} &= \mathbf{r}_1 - \mathbf{r}_2 \end{aligned}\right\} \tag{10.5}$$

The vector differential operators ∇_1 and ∇_2 can be expressed in terms of the quantities ∇_R and ∇_r which correspond to the variables $\mathbf{R}$ and $\mathbf{r}$ respectively by

$$\left.\begin{aligned} \nabla_1 &= \frac{m_1}{m_1 + m_2}\nabla_R + \nabla_r \\ \nabla_2 &= \frac{m_2}{m_1 + m_2}\nabla_R - \nabla_r \end{aligned}\right\} \tag{10.6}$$

Substituting from (10.5) and (10.6) into (10.4) we get

$$\hat{H} = -\frac{\hbar^2}{2M}\nabla_R^2 - \frac{\hbar^2}{2\mu}\nabla_r^2 + V(\mathbf{r}) \tag{10.7}$$

where M is the total mass $(m_1 + m_2)$ and μ is the 'reduced mass', $m_1m_2/(m_1 + m_2)$. The energy eigenvalue equation can now be separated and the eigenfunction written as $U(\mathbf{R})u(\mathbf{r})$ where

$$-\frac{\hbar^2}{2M}\nabla_R^2 U = E_R U \tag{10.8}$$

and

$$\left[-\frac{\hbar^2}{2\mu}\nabla_r^2 + V\right]u = E_r u \tag{10.9}$$

the total energy being equal to $E_R + E_r$. Thus the problem has been separated into an equation that relates only to the centre of mass of the system along with another that describes the behaviour of a particle of mass μ under the influence of a potential $V(\mathbf{r})$. This justifies the procedure used when we obtained the energy levels of the hydrogen atom in chapter 3 (see footnote to p. 53) and when we considered the scattering of neutrons by protons in chapter 9 (problem 9.9).

The separation described here is also possible if the system is not isolated, so long as the external forces can be considered as acting on the centre of mass of the system, because the total potential then has the form $V_R(\mathbf{R}) + V_r(\mathbf{r})$ and equation (10.3) can be separated into two equations, one in $\mathbf{R}$ and the other in $\mathbf{r}$. The motion of the centre of mass can also be separated out in the case

of a similar system containing more than two particles, although the equations describing the internal motion do not then have such a simple form. This explains why the behaviour of the centre of mass of a composite system can be treated without having to consider the detailed behaviour of its component particles. For example, the behaviour of an atom in a gas can be described without taking its internal structure into account, provided the thermal energy is not sufficient to cause electronic excitations. Similarly, the internal structure of the nucleus can be ignored when discussing most of the properties of an atom, while a 'fundamental' particle such as a proton can be considered as a point particle unless very high energy interactions, disturbing its internal structure, are involved.

10.3 Non-interacting particles

We now turn to a case which is more or less the opposite of that just discussed, and consider two particles which may be subject to external forces but which do not interact with each other. The potential can then be written as $V_1(\mathbf{r}_1) + V_2(\mathbf{r}_2)$ and the Hamiltonian becomes

$$-\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + V_1(\mathbf{r}_1) + V_2(\mathbf{r}_2) \tag{10.10}$$

If we now write an energy eigenfunction in the form $u_1(\mathbf{r}_1)u_2(\mathbf{r}_2)$ we can separate the variables $\mathbf{r}_1$ and $\mathbf{r}_2$ and get the following eigenvalue equations

$$-\frac{\hbar^2}{2m_1}\nabla_1^2 u_1 + V_1 u_1 = E_1 u_1 \tag{10.11}$$

$$-\frac{\hbar^2}{2m_2}\nabla_2^2 u_2 + V_2 u_2 = E_2 u_2 \tag{10.12}$$

where E_1 and E_2 are the energies associated with the separate particles and the total energy is

$$E = E_1 + E_2 \tag{10.13}$$

Thus we can apparently treat the particles as completely independent of each other, as would be expected since they do not interact. However, this apparently obvious conclusion is not valid in the particular case where the particles are identical, and we shall discuss this situation in more detail in the next section.

10.4 Indistinguishable particles

Identical particles are often referred to as *indistinguishable* in order to emphasize the fact that they cannot be distinguished by any physical measurement. This implies that an operator representing any physical measurement on the system must remain unchanged if the labels assigned to the individual particles are

interchanged. Thus, if we write the Hamiltonian of two indistinguishable particles as $\hat{H}(1, 2)$, we must have

$$\hat{H}(1, 2) = \hat{H}(2, 1) \tag{10.14}$$

We now define a 'particle interchange operator' $\hat{P}_{12}$ such that, if this operates on any function or operator that depends on the variables describing two particles, it has the effect of interchanging the labels. Thus we have

$$\hat{P}_{12}\hat{H}(1, 2) = \hat{H}(2, 1) = \hat{H}(1, 2) \tag{10.15}$$

using (10.14). Now let $\psi(1, 2)$ be a wavefunction (not necessarily an energy eigenfunction) describing the two particles. The wavefunction is not a physical quantity so need not be invariant when operated on by $\hat{P}_{12}$. However we can write

$$\begin{aligned}\hat{P}_{12}\hat{H}(1, 2)\psi(1, 2) &= \hat{H}(2, 1)\psi(2, 1) \\ &= \hat{H}(1, 2)\hat{P}_{12}\psi(1, 2)\end{aligned} \tag{10.16}$$

Hence

$$[\hat{P}_{12}, \hat{H}(1, 2)] = 0 \tag{10.17}$$

because (10.16) is true whatever the form of ψ. Thus the particle interchange operator and the Hamiltonian commute which means that they are compatible and have a common set of eigenfunctions. Now, because the particles are indistinguishable, not only the Hamiltonian, but any operator representing a physical property of the system must be symmetric with respect to particle interchange and must therefore commute with $\hat{P}_{12}$. Thus whatever measurement is made on the system, the resulting wavefunction will be an eigenfunction of $\hat{P}_{12}$, and no loss of generality is involved if we assume that the wavefunction always has this property.

If $\psi(1, 2)$ is to be an eigenfunction of $\hat{P}_{12}$, it must be a solution of the eigenvalue equation

$$\hat{P}_{12}\psi(1, 2) = p\psi(1, 2) \tag{10.18}$$

where p is the corresponding eigenvalue. But the left-hand side of (10.18) is by definition equal to $\psi(2, 1)$. If we operate on $\psi(2, 1)$ with $\hat{P}_{12}$, we get

$$\begin{aligned}\psi(1, 2) = \hat{P}_{12}\psi(2, 1) &= p\hat{P}_{12}\psi(1, 2) \\ &= p^2\psi(1, 2)\end{aligned}$$

Hence,

$$p = \pm 1 \tag{10.19}$$

and

$$\psi(1, 2) = \pm\psi(2, 1) \tag{10.20}$$

We therefore conclude that any physically acceptable wavefunction representing two identical particles must be either symmetric or antisymmetric with respect

to interchange of the particles. This property is readily extended to a system containing more than two particles when the many-body wavefunction must be either symmetric or antisymmetric with respect to the interchange of any pair of particles. Moreover, we have seen that the operators representing physical measurements all commute with $\hat{P}_{12}$ so that, once a system is in one of these eigenstates, it can never make a transition to the other.

It follows from this that every particle belongs to one of two classes depending on whether the wavefunction representing a number of them is symmetric or antisymmetric with respect to particle exchange. Particles with symmetric wavefunctions are known as *bosons* while those whose wavefunctions are antisymmetric are called *fermions*. This symmetry with respect to particle interchange turns out to be closely connected with the value of the total spin of the particle: bosons always have an integer total-spin quantum number (e.g., the α particle and pion have spin-zero and the deuteron is a spin-one particle) while fermions always have half-integer spin (e.g., the electron, proton, neutron, and neutrino are all spin-half). This simple one-to-one correspondence between the total-spin quantum number and the interchange symmetry is known as the *spin-statistics* theorem, and can be shown to be a necessary consequence of relativistic quantum field theory, and this is briefly discussed in chapter 11.

Non-interacting indistinguishable particles

The results obtained so far in this section apply to any system of indistinguishable particles, whether or not they interact. We can now combine these with the particular properties of non-interacting systems discussed earlier to find the form of the energy eigenfunction of two indistinguishable non-interacting particles. Referring to equations (10.10) we see that, if the particles are indistinguishable, m_1 must equal m_2 and $V_1(\mathbf{r})$ must be the same as $V_2(\mathbf{r})$. Thus equations (10.11) and (10.12) are now identical and have the same set of eigenvalues and eigenfunctions. The total energies of the states with eigenfunctions $u_1(1)u_2(2)$ and $u_1(2)u_2(1)$ are therefore the same, and any linear combination of these is also an eigenfunction with the same eigenvalue. We must therefore form linear combinations of these products which have the appropriate symmetry with respect to particle exchange.

Considering bosons first, the wavefunction must be symmetric with respect to particle exchange so we must have

$$\psi(1,2) = 2^{-1/2}[u_1(1)u_2(2) + u_1(2)u_2(1)] \tag{10.21}$$

where the factor $2^{-1/2}$ ensures that the wavefunction is normalized. In the special case where both particles are associated with the same single-particle state, that is where u_1 is the same as u_2, (10.21) becomes (with a slight change of normalization factor)

$$\psi(1,2) = u_1(1)u_1(2) \tag{10.22}$$

Turning now to fermions, we must form an antisymmetric linear combination of the degenerate functions to get

$$\psi(1,2) = 2^{-1/2}[u_1(1)u_2(2) - u_1(2)u_2(1)] \tag{10.23}$$

and this time the wavefunction in the special case where u_1 is the same as u_2 becomes

$$\begin{aligned}\psi(1,2) &= 2^{-1/2}[u_1(1)u_1(2) - u_1(1)u_1(2)] \\ &= 0\end{aligned}$$

so that the wavefunction now vanishes identically implying that such a state of the system does not exist. This result is known as the *Pauli exclusion principle* and is sometimes expressed by stating that no two fermions can occupy the same state. This can be a useful form of the exclusion principle, but we should realize that if the wavefunction has the form (10.23), a particular particle cannot actually be identified with a particular single-particle function u_1 or u_2. The exclusion principle actually states that each single-particle eigenfunction can be used only once in constructing products, linear combinations of which form the total wavefunctions. We can get away with the simpler form provided we are interested only in the total energy of a system of non-interacting particles. Because of degeneracy, we will get the correct value for the energy if we use simple product wavefunctions of the form $u_1(1)u_2(2)$, and apply the exclusion principle to exclude any products in which both particles are assigned to the same state. We used this procedure in our discussion of the properties of free electrons in solids in chapter 7, where we found that the differences between insulators and metals arise from the fact that all states up to an energy gap are full in the former case, but not the latter. Similar principles can be applied to electrons in atoms to give a qualitative account of many of the chemical properties of the elements and the periodic table. For example, in the ground state of the lithium atom, two of the three electrons occupy the lowest ($1s$) energy level with opposite spin and the third is in the higher-energy $2s$ state, while in the sodium atom, ten electrons fill all the levels with $n = 1$ and $n = 2$, leaving one in the $3s$ level. When all the states with the same n value are full, the electrons in them have comparatively little effect on the physical and chemical properties of the element; it follows that these are largely determined by the one outer electron only and are therefore very similar in the two cases.

If, however, we are interested in properties other than the total energy of a system of non-interacting particles or if we wish to take inter-particle interactions into account, we have to use a wavefunction which is antisymmetric with respect to particle interchange and we must abandon any idea of a particle being associated with a particular state. We shall consider an example of such a system containing two interacting particles when we discuss the helium atom later in this chapter, and we shall return to the conceptual problems associated with the non-separability of identical particles in chapter 13.

10.5 Many-particle systems

The principles discussed in the earlier sections can be readily extended to systems containing more than two particles, but their application to particular cases is often a complex and elaborate process. In this section we shall outline some of the basic ideas to provide an introduction to this topic.

The operators representing the physical properties of a system composed of a number of indistinguishable particles are clearly symmetric with respect to all interchanges of the particle labels, so it follows from the same argument used in the two-particle case that the many-body wavefunction must always be either symmetric or antisymmetric with respect to the exchange of any pair of labels. Thus, for example, the wavefunction representing a system of three bosons must obey the conditions

$$\begin{aligned}\psi(1,2,3) &= \psi(2,1,3) = \psi(3,2,1) = \psi(1,3,2)\\ &= \psi(2,3,1) = \psi(3,1,2)\end{aligned} \tag{10.24}$$

while that representing three fermions must satisfy the following

$$\begin{aligned}\psi(1,2,3) &= -\psi(2,1,3) = -\psi(3,2,1) = -\psi(1,3,2)\\ &= \psi(2,3,1) = \psi(3,1,2)\end{aligned} \tag{10.25}$$

We note that in the latter case the wavefunction is symmetric with respect to a cyclic permutation of the three indices.

The energy eigenvalue equation for non-interacting particles can be separated into as many single-particle equations as there are particles and, assuming these can be solved, the single-particle eigenfunctions $u_1, u_2, \ldots, u_N$ can be formed into products, from which symmetric or antisymmetric linear combinations can be constructed. These can be used as a starting basis in the general case, with the interactions included using perturbation theory.

Bosons

The ground-state eigenfunction of a system of N non-interacting bosons is

$$u_0(1)u_0(2)\ldots u_0(N) \tag{10.26}$$

where u_0 is the single-particle eigenfunction corresponding to the lowest energy eigenvalue, and all the particles are in this state. A gas of non-interacting bosons at zero temperature is described by this wavefunction. At temperatures above absolute zero, the energy eigenfunction is a symmetric linear combination of products of one-electron functions, whose total energy is equal to the internal energy of the gas. This problem can be analysed using statistical methods; the particular form of statistical mechanics in which the exchange symmetry of the wavefunction is preserved is known as *Bose–Einstein statistics*. In general

terms, as the temperature is increased slightly above $T = 0$, some particles are excited into higher states, but most remain in the ground state. However, further rises in temperature give rise to rapid depletion of the ground state, which is completely emptied when a particular temperature known as the Bose condensation temperature (T_0) is reached. A more detailed analysis leads to

$$k_{\mathrm{B}}T_0 = \frac{2\pi\hbar^2}{m}\left[\frac{n}{2.612(2j+1)}\right]^{2/3} \tag{10.27}$$

in the case of n particles of mass m and total spin quantum number j occupying unit volume. The set of atoms occupying the ground state below T_0 constitute what is known as a *Bose condensate*. The properties of this state where a large number of atoms behave as a coherent whole have been subject to considerable investigation. Until the late twentieth century, this was largely focused on the properties of liquid helium; this substance undergoes a transition at about 2.17 K into a 'superfluid state' which exhibits a number of unusual properties, including a complete loss of viscosity. If the atomic mass and liquid density of helium are inserted in (10.27), we get $T_0 = 3\ K$. This indicates that Bose condensation plays an important role in determining the properties of the superfluid state. However, in liquid helium the interatomic energy is of the same order as $k_{\mathrm{B}}T_0$, and condensation and interactions are believed to be of similar importance in this case.

Bose condensation in the absence of interactions was not demonstrated until 1995, when an experiment was carried out on rubidium atoms held in an 'atom trap' constructed from magnets and laser beams. Lasers were also used to bombard the atoms with radiation from all directions. The wavelength of the radiation is chosen so that the photons interact only with atoms that were moving towards them, so slowing them down and producing a temperature of about 10^{-5} K. Rubidium atoms possess spin and therefore have magnetic moments that can be aligned in a field and trapped by appropriately constructed magnetic field gradients, and further cooling is achieved by allowing the more energetic atoms to evaporate from the trap. In this way a few thousand atoms were confined to a volume a few microns in diameter with an interatomic spacing of around 2×10^{-7} m with temperatures as low as $T = 30$ nK. At this density, T_0 is about 170 nK in the case of rubidium. As this separation is about 400 times larger than the distance between atoms in solid rubidium, the interactions between the atoms are negligibly small.

The evidence for the creation of a Bose condensate in this way is illustrated in figure 10.1. After compressing and cooling the atoms, the confining field is removed and the spatial distribution of the atoms is recorded a short time later. If the atoms are in thermal motion, they move randomly during this time and are seen to be well spread out on the scale of the diagram—as is the case at $T = 200$ nK. However, if they have condensed into their ground state, the spread is due only to the quantum uncertainty in the momentum, which is much smaller. As a result, the distributions at $T = 100$ nK and $T = 30$ nK contain

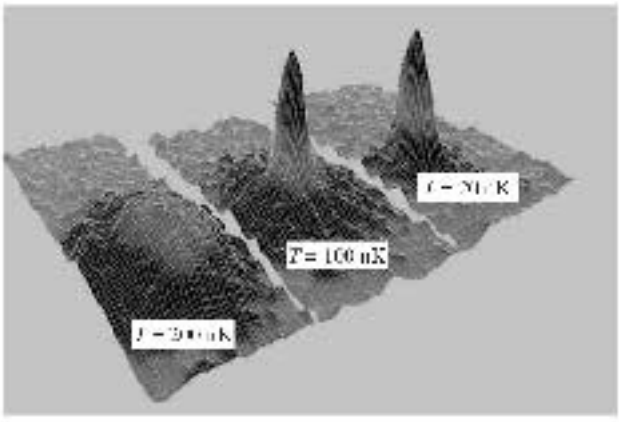


Figure 10.1. The distribution of atomic velocities in a collection of rubidium atoms held in a trap for three temperatures that span the Bose condensation temperature of 170 nK. Zero velocity corresponds to the centre of the peak in each case. The condensate is absent at 200 nK, well established at 100 nK and very nearly complete at 20 nK. (As illustrated for the JILA Bose–Einstein condensate group in 1995 by M. R. Matthews, and reproduced by permission.)

the narrow peaks shown. The fact that this is a thermodynamic phase transition is confirmed by the very rapid growth in the condensate as the temperature is lowered through the transition temperature. In recent years, the detailed properties of Bose condensates have been subject to considerable experimental and theoretical investigation.

Fermions

In the case of non-interacting fermions, antisymmetric linear combinations have to be formed from the products of single-particle eigenfunctions. The general form of these in the case of a system of N particles, where the single-particle eigenfunctions are $u_1, u_2, \ldots, u_N$, is

$$\psi = \frac{1}{(N!)^{1/2}} \begin{vmatrix} u_1(1) & u_2(1) & \cdots & u_N(1) \\ u_1(2) & u_2(2) & \cdots & u_N(2) \\ \vdots & \vdots & & \vdots \\ u_1(N) & u_2(N) & \cdots & u_N(N) \end{vmatrix} \tag{10.28}$$

Such an expression is known as a *Slater determinant*. We can see that this fulfils the required condition of antisymmetry using standard properties of determinants:

an exchange of the labels on two particles is equivalent to exchanging two rows of the determinant which leaves its magnitude unchanged, but reverses its sign. We also note that the expansion of an $N \times N$ determinant contains $N!$ terms which leads to the normalization factor of $(N!)^{-1/2}$, assuming the one-electron orbitals are orthonormal. The Pauli exclusion principle also follows from (10.28): if two or more particles are associated with the same single-particle eigenfunction, two or more columns of the determinant are identical and the whole wavefunction vanishes.

The form of statistical mechanics applying to a gas of non-interacting fermions is known as *Fermi–Dirac* statistics. At zero temperature the wavefunction has the form (10.28) where the single-particle eigenfunctions are those with smallest eigenvalues, while at higher temperatures states with higher energy are involved. We saw an example of a system obeying Fermi–Dirac statistics when we discussed the behaviour of electrons in metals in chapter 7; another example is a gas of atoms of the isotope ^{3}He whose nucleus has a total spin quantum number of one-half. As the electrons have zero total orbital angular momentum and zero spin in the ground state, the total wavefunction must be asymmetric with respect to exchange of pairs of hydrogen atoms and such a gas must obey Fermi–Dirac statistics. As a result the low-temperature properties of ^{3}He are quite different from those of normal helium whose nucleus, ^{4}He, has zero spin. For example, ^{4}He exhibits superfluidity below a temperature of about 2 K, but ^{3}He remains a normal liquid down to temperatures less than 0.01 K. At high temperatures, however, the properties of the two gases are very similar; and this can also be shown to be in agreement with the results of quantum statistical mechanics.

The case of interacting fermions has considerable practical importance because it represents the situation applying in many-electron atoms and molecules as well as in nuclei and in the case of electrons in solids. Exact solutions are rarely possible, but various approximations have been developed. One of the most important of these is known as the self-consistent field or Hartree-Fock method. In this approximation, the single-particle eigenfunctions are calculated assuming that each particle is subject to a potential equivalent to the average of its interaction with all the others. This interaction potential is taken to be that from a continuous charge distribution whose magnitude is proportional to the sum of the squared moduli of some approximate set of one-electron eigenfunctions, along with an exchange term analogous to that discussed later in the case of the helium atom (10.36). The resulting eigenvalue equations are then solved to give new one-electron eigenfunctions which, in turn, are used to construct new potentials. This iterative process is continued until no further significant changes occur, when the system is said to be 'self-consistent'. It can be shown that this procedure is equivalent to using the variational principle to obtain the best possible representation of the many-particle eigenfunction in the form of a single determinant such as (10.28). More accurate results can be obtained if the eigenfunction is represented by a linear combination of determinants, each of

which is constructed from a different set of one-electron functions. Nowadays, such calculations are quite straightforward for all but the heaviest atoms and for moderately sized molecules. Computations of this kind now form part of the routine armoury of the theoretical chemist and are particularly powerful for the study of short-lived chemical species, which often cannot be studied experimentally.

10.6 The helium atom

The helium atom consists of two electrons and a double charged positive nucleus. Throughout our discussion we shall make the approximation that the mass of the nucleus is infinitely greater than that of an electron so that the problem can be treated as that of two electrons moving in a potential. For the moment we shall also ignore spin–orbit coupling so that the Hamiltonian of the system is

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla_1^2 - \frac{\hbar^2}{2m}\nabla_2^2 - \frac{2e^2}{4\pi\varepsilon_0 r_1} - \frac{2e^2}{4\pi\varepsilon_0 r_2} + \frac{e^2}{4\pi\varepsilon_0 r_{12}} \tag{10.29}$$

where $r_{12} \equiv |\mathbf{r}_1 - \mathbf{r}_2|$ is the separation of the two electrons. We note that, as expected, $\hat{H}$ is symmetric with respect to the interchange of the labels on the identical electrons. In the discussion that follows we first ignore the last term in (10.29), which represents the inter-electronic interaction, so that we can consider the problem as a non-interacting one. The effects of the interaction term are then treated by perturbation theory and, finally, spin–orbit interactions will be considered.

We saw earlier that in the non-interacting limit, the energy eigenfunctions can be expressed as products of single-particle eigenfunctions, which in general depend on both the particle position and its spin. In the case of helium it follows that, if the interaction is ignored, the spatial parts of the single-particle eigenfunctions are solutions to the equation

$$\left(-\frac{\hbar^2}{2m}\nabla_1^2 - \frac{2e^2}{4\pi\varepsilon_0 r_1}\right) u = Eu \tag{10.30}$$

and are therefore the same as the hydrogenic eigenfunctions described in chapter 3 and referred to several times since. It will also be important to include the spin part of the wavefunction and we represent this by α or β depending on whether the z component of spin is positive or negative respectively. If we now consider states where the spatial parts of the single-particle eigenfunctions are u_1 and u_2, and if we include all possible values of the spin, we obtain the following eight products, all of which are energy eigenfunctions with the same eigenvalue in the

non-interacting limit:

$$\left.\begin{array}{ll} u_1(\mathbf{r}_1)\alpha(1)u_2(\mathbf{r}_2)\alpha(2) & u_2(\mathbf{r}_1)\alpha(1)u_1(\mathbf{r}_2)\alpha(2) \\ u_1(\mathbf{r}_1)\alpha(1)u_2(\mathbf{r}_2)\beta(2) & u_2(\mathbf{r}_1)\beta(1)u_1(\mathbf{r}_2)\alpha(2) \\ u_1(\mathbf{r}_1)\beta(1)u_2(\mathbf{r}_2)\alpha(2) & u_2(\mathbf{r}_1)\alpha(1)u_1(\mathbf{r}_2)\beta(2) \\ u_1(\mathbf{r}_1)\beta(1)u_2(\mathbf{r}_2)\beta(2) & u_2(\mathbf{r}_1)\beta(1)u_1(\mathbf{r}_2)\beta(2) \end{array}\right\} \tag{10.31}$$

We construct antisymmetric functions by taking appropriate linear combinations of the products listed in (10.31). This can be done in a number of ways, but it will turn out to be an advantage if they are each expressed as a product of a spatially dependent and a spin-dependent part; in this case, if the total wavefunction is to be antisymmetric, either the spin-dependent part must be antisymmetric and the spatial part symmetric or vice versa. The four functions which can be constructed consistently with these requirements are then:

$$\left.\begin{array}{l} \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) - u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)]\alpha(1)\alpha(2) \\ \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) - u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)]\beta(1)\beta(2) \\ \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) - u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)]\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) + \alpha(2)\beta(1)] \\ \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) + u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)]\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)] \end{array}\right\} \tag{10.32}$$

It should be noted that the spin parts of the functions are eigenfunctions of the operators representing the total spin of the two particles and of their total z component (cf. the discussion of the coupling of angular momentum in sections 6.5 and 6.6): the first three functions listed have their total-spin quantum number S equal to 1 and the quantum number associated with the z component m_s has the values 1, -1 and 0 respectively; the fourth function has zero spin, so that $S = m_s = 0$. Thus in the first three states the electron spins are aligned as nearly parallel as is allowed by the quantization rules for angular momentum, while in the fourth they are antiparallel. We also note that in the special case where the spatial parts of the single-particle functions are identical, that is, where $u_1 \equiv u_2$, the only state allowed by the exclusion principle is the one with $S = 0$ and the others are identically zero as expected.

We shall now consider how the energies of these states are affected by the inclusion of the inter-electronic electrostatic interaction and by spin–orbit coupling.

Inter-electronic interactions

We consider first the effects of the electrostatic interactions between the electrons. We can ignore the spin parts of (10.32) for the moment, and apply perturbation

theory to two degenerate states whose unperturbed eigenfunctions are

$$\left.\begin{aligned} v_{01} &= \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) - u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)] \\ \text{and} \qquad v_{02} &= \frac{1}{\sqrt{2}}[u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) + u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)] \end{aligned}\right\} \tag{10.33}$$

where the perturbation is

$$H' = \frac{e^2}{4\pi\varepsilon_0 r_{12}} \tag{10.34}$$

Following the standard procedure described in chapter 7 we form the matrix elements of $\hat{H}'$. Using (10.33) the off-diagonal element $H'_{12} = \int v_{01}^* H' v_{02} d\tau_1 d\tau_2$ can be expressed as four integrals involving u_1 and u_2. These integrals form two pairs that have identical magnitude and opposite sign, so $H'_{12} = 0$. There is therefore no mixing of the two states and v_{01} and v_{02} are the appropriate zero-order eigenfunctions to be used when calculating the effects of the perturbation. The first-order changes (E_{11} and E_{12}) to the energies of the two states are then equal to the diagonal elements of the perturbation matrix and we have

$$\left.\begin{aligned} E_{11} &= H'_{11} = C - X \\ E_{12} &= H'_{22} = C + X \end{aligned}\right\} \tag{10.35}$$

where

$$\left.\begin{aligned} C &= \iint u_1^*(\mathbf{r}_1)u_2^*(\mathbf{r}_2)\frac{e^2}{4\pi\varepsilon_0 r_{12}}u_1(\mathbf{r}_1)u_2(\mathbf{r}_2)\,d\tau_1\,d\tau_2 \\ \text{and} \qquad X &= \iint u_1^*(\mathbf{r}_1)u_2^*(\mathbf{r}_2)\frac{e^2}{4\pi\varepsilon_0 r_{12}}u_1(\mathbf{r}_2)u_2(\mathbf{r}_1)\,d\tau_1\,d\tau_2 \end{aligned}\right\} \tag{10.36}$$

The integral C is sometimes referred to as the 'Coulomb energy' because it is equivalent to that of the classical electrostatic interaction between two continuous charge distributions whose densities are $|u_1|^2$ and $|u_2|^2$ respectively, while X is often referred to as the 'exchange energy' because the integrand contains terms where the numbers labelling the electrons have been 'exchanged'. Very little physical significance should be attached to these designations, however, as the integrals arise simply as a result of applying perturbation theory to the system, and the only physically significant quantity being calculated is the total energy whose value is equal to either the sum or difference of the two terms.

The quantities C and X can be evaluated given the form of the one-electron eigenfunctions u_1 and u_2. The procedure is reasonably straightforward in principle, but rather complex and tedious in detail. Accordingly we shall not describe such calculations here, but concentrate on the qualitative significance of the expressions for the energy eigenvalues, although we shall briefly discuss the application of equations (10.36) to the calculation of the ground-state energy towards the end of this section.

We first note that, whereas there are four degenerate eigenfunctions (10.32) in the absence of a perturbation, the inclusion of the electrostatic interaction has split this system into a 'triplet' of three degenerate states all of which have a total-spin quantum number equal to one, along with a spin-zero 'singlet' state. Thus the energies of the states where the spins are parallel are different from those where the spins are antiparallel and the system behaves as though the spins were strongly coupled. It is important to remember that this coupling results from the requirement that the wavefunction be antisymmetric and is quite independent of the interaction between the magnetic dipoles associated with the spins which is very much weaker. It turns out that the exchange integral X is generally positive so that the triplet states have lower energy than the corresponding singlets. Nevertheless, the ground state of helium is a singlet, because in this case both electrons occupy the same $(1s)$ orbital and no corresponding triplet state exists. Another type of physical system that exhibits a similar coupling of spins is a ferromagnet such as iron: in this case, however, the exchange interaction between electrons associated with neighbouring atoms happens to lead to a negative X and hence a triplet ground state. The magnetic moments associated with the atomic spins therefore all point in the same direction, leading to a large magnetic moment overall.

Returning to the helium atom, the coupling of the spins via the exchange energy means that the spin–orbit interaction is between the total spin and the total orbital angular momentum, whose magnitude is determined by the quantum numbers of the single-particle functions u_1 and u_2. Given these quantum numbers, the splitting due to spin–orbit coupling can be calculated by the methods discussed in chapter 6, as can the response of the system to weak and strong magnetic fields. Helium is therefore an example of Russell–Saunders coupling which was discussed briefly at the end of chapter 6 where we mentioned that it applied most usefully to atoms of low atomic number. In the case of heavy atoms, the magnetic interaction between the spin and orbital angular momenta of the individual electrons turns out to be stronger than the exchange interaction and we get j–j coupling.

The fact that the spatial part of a singlet wavefunction is symmetric, while that of a triplet is antisymmetric, means that electric dipole transitions between any members of these two sets of states are forbidden. This is because the electric dipole operator (cf. chapter 8) is a function of the electron positions only and, like all physical operators, is symmetric with respect to particle interchange so that the matrix elements connecting the singlet and triplet states are of the form

$$\begin{aligned}\iint \psi_s^*(1,2)\,\hat{Q}_s(1,2)\psi_a(1,2)\,d\tau_1\,d\tau_2 \\ = -\iint \psi_s^*(2,1)\,\hat{Q}_s(2,1)\psi_a(2,1)\,d\tau_1\,d\tau_2 \end{aligned} \tag{10.37}$$

where the subscripts s and a signify symmetry and antisymmetry respectively and $\hat{Q}_s$ is any symmetric operator. But the labels on the variables of integration

have no significance, so the integrals on each side of (10.37) are identical, which can be true only if they both vanish. The same is true for any perturbation that is a function of either the electron positions or of the spins only. It follows that singlet to triplet transitions, and vice versa, occur only as a result of collisions between atoms, where the property conserved is the total antisymmetry of the wavefunction representing all the electrons in both atoms, rather than that of each atom separately. Because transitions between singlet and triplet states occur so rarely, they were not observed at all in the early days of spectroscopy and, at one time, helium was thought to be a mixture of two gases: 'parahelium' with the singlet spectrum and 'orthohelium' with that of the triplet.

These properties also underlie the principles of operation of the helium–neon laser. When an electric discharge is passed through helium gas, it causes many of the atoms to be ionized. When they subsequently recombine, there is an appreciable probability of some of the atoms being in one of the excited triplet states. These then decay, emitting photons, until they reach the lowest energy triplet state where they remain because a further transition to the singlet ground state is forbidden. It turns out that the excitation energy of the lowest triplet state of helium is close to that of one of the excited states of neon so that in a mixture of the two gases there is an appreciable probability of neon atoms being excited by collisions with appropriate helium atoms. If the partial pressures of the gases in such a mixture are right, an inverted population can be generated in which more neon atoms are in this excited state than are in another state which has lower energy. Radiation whose frequency matches the energy difference between the two neon states can therefore be amplified by stimulated emission and laser action ensues. This can be maintained continuously by passing a suitable discharge through the mixture, thereby replenishing the number of triplet helium atoms.

We shall complete our discussion of the helium atom by considering some of the more quantitative features of the energy levels. Figure 10.2 shows an energy-level diagram in which the experimentally measured energies of the states are referred to an origin that corresponds to the energy required to just remove one electron from the atom, leaving the other in the ground state of the remaining He^+ ion. The energy levels of the hydrogen atom derived in chapter 3 are also included for comparison. We see that, on this scale, the ground-state energy of helium is very much lower than that of hydrogen and that this is also true, to a lesser extent, of the first and second excited states. However, states with high values of the principal quantum number n (i.e. greater than about five) have energies that are very little different from the corresponding hydrogen-atom values, being nearly independent of the orbital-angular-momentum quantum number l. This last point is consistent with the fact that helium, like hydrogen, exhibits a first-order Stark effect (cf. section 7.2).

We can account for all these features using the theory developed earlier. The first point to note is that all states where both electrons are assigned to excited orbitals turn out to have such a high energy that they cannot exist as bound states. All the bound states of the helium atom therefore involve linear combinations

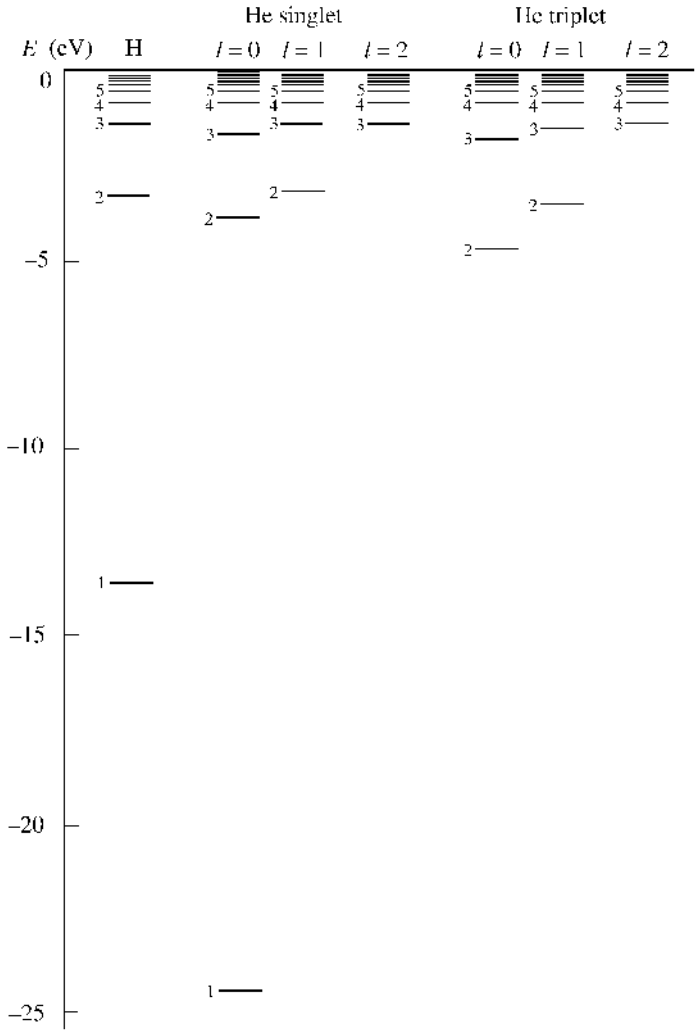


Figure 10.2. The energies of the singlet and triplet states of the helium atom compared with the energy levels of the hydrogen atom. The number beside each energy level is the principal quantum number n. Each triplet state is actually three closely spaced levels split by the spin–orbit interaction. Zero energy corresponds to a singly ionized He atom.

of products in which one of the one-electron functions is the lowest energy ($1s$) orbital. Accordingly, the quantity previously referred to as the principal quantum number n denotes the orbital other than the $1s$ that is involved in the description of the state.

We first consider highly excited states and note that the one-electron orbitals

with high values of n have appreciable magnitude only at large distances from the nucleus, while the $1s$ eigenfunction is significant only at small values of r (cf. figure 3.4). If we now examine the expressions for the Coulomb and exchange energies (10.36) we see that, for a highly excited state, the exchange energy is very small because u_1 and u_2 occupy almost entirely different regions of space. The singlet and triplet states can therefore be treated as degenerate, which means that the energy can be estimated by assigning one electron to each of the orbitals u_1 and u_2 in the same way as in the non-interacting case discussed earlier. The total energy of such a state would then be equal to the sum of the energies of the ground and appropriately excited states of the He^+ ion, along with the Coulomb energy which represents the mean electrostatic interaction between electrons assigned to these two orbitals. This result still has the limitations of first-order perturbation theory and a better approximation is obtained by assuming that the particles can be independently assigned to each single-particle orbital but that the form of each orbital is affected by the presence of the other electron. Thus the outer electron moves in a potential due to the nucleus and the charge distribution of the spherically symmetric inner orbital, which is equivalent to that of a single positive charge $(+e)$ at the nucleus. The outer electron therefore has an energy close to that of the corresponding state in hydrogen, while the inner electron has an energy equivalent to that of the $1s$ orbital of a He^+ ion, because the field inside the spherically symmetric charge distribution of the outer orbital is zero. The total energy of such an excited state is therefore close to the sum of the ground-state energy of a He^+ ion and that of the appropriately excited state of the hydrogen atom, which is just the result obtained experimentally and described in the previous paragraph (cf. figure 10.1).

Turning our attention to the ground state of the helium atom, in this case both electrons are associated with the $1s$ orbital, so the exchange term again vanishes (this time because $u_1 \equiv u_2$) and the effect of the Coulomb term is that each electron, to some extent, screens the other from the full potential of the doubly charged nucleus. Evaluation of the Coulomb integral for the ground state yields a value of 34.0 eV which, when combined with the unperturbed energy of -54.4 eV, produces a value of -20.4 eV for the total, in good agreement with the experimental result of -24.8 eV, given the limitations of first-order perturbation theory in this context.

Finally, we consider states, other than the ground state, that are not highly excited. We can now no longer assume that the exchange energy is negligible, and the singlet and triplet states are therefore expected to have different energies, as is observed experimentally. Moreover, the effective potential experienced by the electrons is now significantly different from the Coulomb form, so states whose orbitals have the same values of n, but different l, are no longer degenerate. We see in figure 10.2 that both these effects are most pronounced for states constructed from an orbital with $n = 2$ along with the $1s$ orbital, and become progressively smaller as n increases until, for n greater than about five, the states can be considered as 'highly excited' in the sense described earlier.

10.7 Scattering of identical particles

We close this chapter by considering the problem of scattering where the particles in the incident beam are identical to that constituting the scatterer; an example to which we shall return shortly is the scattering of alpha particles by the nuclei of ^{4}He atoms. We note that in a scattering problem there are no external forces, so we can separate the relative motion from that of the centre of mass using the procedure discussed previously for isolated systems. Neglecting spin for the present, the energy eigenfunctions can be written as

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = U(\mathbf{R})u(\mathbf{r}) \tag{10.38}$$

where—using (10.5) and remembering that $m_1 = m_2$—$\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$ and $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. Clearly $U(\mathbf{R})$ is completely symmetric with respect to particle exchange, so the symmetry of the wavefunction is determined by that of $u(\mathbf{r})$. Moreover, particle interchange is equivalent to a reversal of the sign of $\mathbf{r}$ so exchange symmetry is equivalent to the parity in this case.

In chapter 9 (cf. equations (9.57) to (9.60)) we showed that the energy eigenfunction in a scattering problem could be written in the form

$$V^{-1/2}\left[e^{ikz} + \frac{1}{kr}e^{ikr}f(\theta)\right] \tag{10.39}$$

where

$$f(\theta) = \sum_{l=0}^{\infty}(2l+1)e^{i\delta_l}\sin\delta_l P_l(\cos\theta) \tag{10.40}$$

Equation (10.39) does not have a definite parity so the correct eigenfunction in the case of identical particles must be equal to a linear combination of this expression with a similar one whose sign of $\mathbf{r}$ is reversed. That is,

$$u(\mathbf{r}) = (2V)^{-1/2}\left\{[e^{ikz} \pm e^{-ikz}] + \frac{1}{kr}e^{ikr}[f(\theta) \pm f(\pi-\theta)]\right\} \tag{10.41}$$

where the positive and negative signs apply when $u(\mathbf{r})$ is symmetric and antisymmetric respectively. In the absence of scattering, the wavefunction is proportional to the first term in square brackets in (10.41), so it follows by an argument similar to that leading to equations (9.61) and (9.62), that the differential cross section $\sigma(\theta)$ is given by

$$\sigma(\theta) = |f(\theta) \pm f(\pi-\theta)|^2 \tag{10.42}$$

We know from chapter 3 that

$$P_l(\cos\theta) = (-1)^l P_l(\cos(\pi-\theta))$$

so it follows from (10.41) and (10.42) that the cross section is made up from partial waves with only even values of l in the symmetric case and with only odd

values in the asymmetric case. It also follows from (10.42) that the contributions to the total cross section from those partial waves that have non-zero amplitude are four times what they would be if the particles were distinguishable. Half of this factor of four arises because a recoiling target particle cannot be distinguished from one scattered out of the incident beam, but the remaining factor of two has no such simple cause. All these results have been confirmed experimentally by scattering experiments involving spin-zero particles: the scattering of α particles by ^{4}He atoms, for example, departs from the predictions of the Rutherford scattering formula (9.42) in a manner that can be quantitatively accounted for on this basis.

When the particles have non-zero spin, the overall symmetry of the wavefunction is determined by that of both the spatial and the spin-dependent parts. We shall not consider the general case here, but confine our discussion to spin-half fermions where the total wavefunction must be antisymmetric. It follows that all partial waves contribute to the scattering, but for those with even l the spin part must be antisymmetric and therefore correspond to a singlet state with zero total spin, while for those with odd l the spin function must be symmetric, corresponding to a spin-one triplet state. Thus if, for example, the energy of the incident particle were such that the scattering is s-wave in character, we can conclude that scattering will occur only when the particle spins correspond to a singlet state. As, on average, only one-quarter of the particle pairs is in this state, the total cross section is reduced by a factor of four to be the same as if the particles were distinguishable. After such a scattering event, therefore, the incident and target particles move off in a state where both the total orbital angular momentum and the total spin are zero. An example of the application of this principle is the scattering of protons by hydrogen nuclei at an energy large enough to ensure that the scattering is predominantly due to the nuclear force rather than the Coulomb interaction, but small enough to ensure that the contribution from partial waves with non-zero values of l is negligible. As will be discussed in chapter 13, measurements of the properties of such pairs of protons can be used to compare the predictions of quantum mechanics with those of 'hidden variable' theories.

Problems

10.1 Show that for a many-particle system, subject to no external forces, the total energy and total momentum can always be measured compatibly, but that, if the particles interact, the individual momenta cannot be measured compatibly with the total energy.

10.2 Two particles of masses m_1 and m_2 move in one dimension and are not subject to any external forces. The potential energy of interaction between the particles is given by

$$V = 0 \quad (|x_{12}| \leqslant a); \qquad V = \infty \quad (|x_{12}| > a)$$

where x_{12} represents the particle separation. Obtain expressions for the energy eigenvalues and eigenfunctions of this system if its total momentum is P.

10.3 Repeat the calculation in problem 10.2 for the case where the two particles have the same mass

m and are (i) indistinguishable spin-zero bosons and (ii) indistinguishable spin-half fermions.

10.4 Show that if $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ is the wavefunction representing a system of N indistinguishable particles, then the probability of finding any one of these in the element $d\tau$ around the point $\mathbf{r}$ is given by $P_1(\mathbf{r})\, d\tau$ where

$$P_1(\mathbf{r}) = \int \ldots \int |\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)|^2 \, d\tau_2 \ldots d\tau_N$$

Show that in the case of two non-interacting particles (either bosons or fermions)

$$P_1(\mathbf{r}) = \tfrac{1}{2}(|u_1(\mathbf{r})|^2 + |u_2(\mathbf{r})|^2)$$

where u_1 and u_2 are appropriate single-particle eigenfunctions.

10.5 Two indistinguishable non-interacting spin-half particles move in an infinite-sided one-dimensional potential well. Obtain expressions for the energy eigenvalues and eigenfunctions of the ground and first excited states and use these to calculate $P_1(x)$—as defined in problem 10.4—in each case.

Show that in those of these states where the total spin is zero, there is generally a finite probability of finding the two particles at the same point in space, but that this probability is zero if the total spin is one.

10.6 Show that if the potential described in problem 10.5 has width Na and contains a large number N of non-interacting fermions, the total ground-state energy is approximately equal to $(\pi^2\hbar^2/24ma^2)N$.

10.7 Using the one-dimensional model described in problem 10.6, estimate the average energy per particle for (i) a free-electron gas and (ii) a gas of ^{3}He atoms assuming a mean linear density of 4×10^9 particles per metre in each case.

Compare your answer to (ii) with a similar estimate of the ground-state energy per atom of ^{4}He gas. Above about what temperature might you expect these gases to have similar properties?

10.8 A bound system consisting of two neutrons is almost, but not quite, stable. Estimate the energy of the virtual energy level of this system given that the low-energy s-wave scattering cross section of neutrons by neutrons is about 60×10^{-28} m^2.

Hint: cf. problem 9.9.

Chapter 11

Relativity and quantum mechanics

The early twentieth century saw two major revolutions in the way physicists understand the world. One was quantum mechanics itself and the other was the theory of relativity. Important results also emerged when these two ideas were brought together; some have been referred to in earlier chapters—in particular the fact that fundamental particles such as electrons have intrinsic angular momentum (spin) was stated to be a relativistic effect.

A full understanding of relativistic quantum mechanics is well outside the scope of this book, but many of the most important results can be understood at this level and these will be discussed in this chapter. After a short summary of the main results of special relativity, we show how combining this with the time-dependent Schrödinger equation leads to a new wave equation known as the Dirac equation. We show how the Dirac equation requires particles such as electrons to have intrinsic angular momentum (spin) and we explore some of its other consequences. The chapter concludes with an outline of some more advanced ideas known as quantum field theory.

11.1 Basic results in special relativity

Special relativity modifies classical (i.e. non-quantum) kinematics and dynamics to encompass phenomena that are exhibited strongly when particles move at speeds comparable to the speed of light. We will assume that the reader is familiar with the main results of special relativity, and this section will be restricted to a summary of those needed for our later discussion.

The kinematics are governed by the Lorentz transformation which relates the position and time coordinates (x, y, z, t) of an event observed in one inertial frame of reference to those (x', y', z', t') observed from another moving at constant speed v in the x direction relative to the first. We have

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}} \qquad y' = y \qquad z' = z \qquad t' = \frac{t - (v/c^2)x}{\sqrt{1 - v^2/c^2}} \tag{11.1}$$

where c is the speed of light. The momentum $\mathbf{p}$ and energy E of a particle also transform by a Lorentz transformation where x, y, z and ct in (11.1) are replaced by p_x, p_y, p_z and E/c respectively, the energy including the 'rest-mass' energy mc^2. The energy and momentum are then related by the principal equation governing the dynamics of a free particle:

$$E^2 = p^2c^2 + m^2c^4 \tag{11.2}$$

Equation (11.2) is an example of a *Lorentz invariant*—i.e. it has the same form in all inertial frames of reference, as can be verified by applying the Lorentz transformation to the components of $\mathbf{p}$ and E. Also, if we put $E = mc^2 + \epsilon$, the non-relativistic limit is when $\epsilon \ll mc^2$, in which case (11.2) reduces to $\epsilon = p^2/2m$.

We now consider the case where the particle is not free, but subject to an electromagnetic field. This field can, in turn, be represented by a scalar potential, $\phi(\mathbf{r})$ and vector potential, $\mathbf{A}(\mathbf{r})$ where the electric field is $\mathcal{E} = -\nabla\phi + \partial\mathbf{A}/\partial t$ and the magnetic field is $\mathbf{B} = \nabla \times \mathbf{A}$. Equation (11.2) then becomes for a particle of charge q

$$(E - q\phi)^2 = (\mathbf{p} - q\mathbf{A})^2c^2 + m^2c^4 \tag{11.3}$$

11.2 The Dirac equation

A successful relativistic wave equation for a particle such as an electron was first obtained by P. A. M. Dirac in 1928. In the same way as the Schrödinger equation cannot be derived from classical mechanics because it is essentially new physics, any relativistic equation can only be guessed at by a process of induction, and its truth or otherwise must be established by testing its consequences against experiment. Following Dirac, we start this process by considering the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\psi = \hat{H}\psi \tag{11.4}$$

As in the non-relativistic case, (cf. chapter 2) we will begin by considering the case of a free particle—i.e. $\mathbf{A} = \phi = 0$. Following the principles of postulate 3 in chapter 4, we assume that the energy operator $\hat{H}$ can be expressed in terms of the momentum operator $\hat{\mathbf{P}}$ in the same way as E is related to $\mathbf{p}$ in the classical limit. Hence, using (11.2) and (11.4),

$$i\hbar\frac{\partial\psi}{\partial t} = \sqrt{[\hat{P}^2c^2 + m^2c^4]}\ \psi = 0 \tag{11.5}$$

We are now faced with the problem of an operator that is a square root of another operator. There is no definite prescription for handling this, but we do know that, in order to preserve Lorentz invariance, positional coordinates and time must appear in a similar way in any relativistic theory. So if we are to make

the standard replacement $\hat{P}_x = -i\hbar\frac{\partial}{\partial x}$ etc., the second term of (11.5) should be linear in these quantities. Applying this principle and following Dirac we write

$$i\hbar\frac{\partial}{\partial t}\psi = [c\alpha_1\hat{P}_x + c\alpha_2\hat{P}_y + c\alpha_3\hat{P}_z + \beta mc^2]\psi \tag{11.6}$$

where the α_i and β are dimensionless quantities that are independent of position and time.

For (11.5) and (11.6) to be consistent, the squares of the operators on the right-hand sides of these equations should be equivalent. That is

$$[c\alpha_1\hat{P}_x + c\alpha_2\hat{P}_y + c\alpha_3\hat{P}_z + \beta mc^2]^2 = c^2\hat{P}^2 + m^2c^4 \tag{11.7}$$

Multiplying out the left-hand side of (11.7) and equating corresponding terms leads to

$$\left.\begin{aligned} \alpha_1^2 = \alpha_2^2 = \alpha_3^2 = 1 \\ \alpha_1\alpha_2 + \alpha_2\alpha_1 = \alpha_2\alpha_3 + \alpha_3\alpha_2 = \alpha_3\alpha_1 + \alpha_1\alpha_3 = 0 \\ \alpha_1\beta + \beta\alpha_1 = \alpha_2\beta + \beta\alpha_2 = \alpha_3\beta + \beta\alpha_3 = 0 \end{aligned}\right\} \tag{11.8}$$

This is obviously not possible if α_i and β are scalar numbers, and Dirac showed that the simplest form of α_i and β was a set of 4×4 matrices:

$$\alpha_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \alpha_2 = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix}$$

$$\alpha_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \qquad \beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{11.9}$$

The reader should check that these expressions have the properties set out in (11.8).

We note the appearance of the 2×2 Pauli spin matrices within these 4×4 matrices. These are given in chapter 6 as

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{11.10}$$

We can therefore re-write (11.9) in terms of these and the 2×2 unit matrix (represented by I):

$$\alpha_j = \begin{bmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{bmatrix} \qquad \beta = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \qquad j = 1, 2, 3 \tag{11.11}$$

where $\sigma_1 \equiv \sigma_x$ etc. The following properties of the Pauli spin matrices, which can be easily proved by direct substitution, will be used shortly

$$\left.\begin{aligned} \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I \\ \sigma_x\sigma_y = -\sigma_y\sigma_x = i\sigma_z \end{aligned}\right\} \qquad (11.12)$$

$$\sigma_x\sigma_y = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \qquad \sigma_y\sigma_x = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}$$

Similar results hold for cyclic permutations of the Cartesian coordinates.

The fact that the Dirac equation is a matrix equation implies that ψ is a vector formed out of four functions of position and time:

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{bmatrix} \equiv \begin{bmatrix} \psi_+ \\ \psi_- \end{bmatrix} \qquad (11.13)$$

where ψ_+ and ψ_- are two-component vectors. The Dirac equation (11.6) can then be written in the form

$$\left.\begin{aligned} (\boldsymbol{\sigma}\cdot\mathbf{p})c\psi_- + mc^2\psi_+ = i\hbar\frac{\partial\psi_+}{\partial t} \\ (\boldsymbol{\sigma}\cdot\mathbf{p})c\psi_+ - mc^2\psi_- = i\hbar\frac{\partial\psi_-}{\partial t} \end{aligned}\right\} \qquad (11.14)$$

We can separate out the time dependence in the same way as we did in the non-relativistic case by putting $\psi = u\exp(-iEt/\hbar)$ to get

$$\left.\begin{aligned} (\boldsymbol{\sigma}\cdot\mathbf{p})cu_- + mc^2u_+ = Eu_+ \\ (\boldsymbol{\sigma}\cdot\mathbf{p})cu_+ - mc^2u_- = Eu_- \end{aligned}\right\} \qquad (11.15)$$

We note again that all the terms in (11.15) are second-order matrices so that each of these equations actually represents two equations. We can use the second of (11.15) to express u_- in terms of u_+:

$$u_- = \frac{c}{mc^2 + E}(\boldsymbol{\sigma}\cdot\mathbf{p})u_+ \qquad (11.16)$$

We substitute this into the first of (11.15) to get

$$(\boldsymbol{\sigma}\cdot\mathbf{p})^2c^2u_+ = (E - mc^2)(E + mc^2)u_+ \qquad (11.17)$$

Expanding the first term on the left-hand side in Cartesian coordinates and re-arranging, we get

$$[\sigma_x^2\hat{P}_x^2 + (\sigma_x\sigma_y + \sigma_y\sigma_x)\hat{P}_x\hat{P}_y + \circ]c^2u_+ = (E^2 - m^2c^4)u_+ \qquad (11.18)$$

where the symbol $+_\circ$ implies repeating the previous expression twice, with cyclic permutation of the Cartesian coordinates. Using (11.12), we can rewrite (11.18) as

$$\hat{P}^2c^2u_+ = (E^2 - m^2c^4)u_+ \tag{11.19}$$

We can solve (11.19) in the representation where $\hat{\mathbf{P}} = -i\hbar\nabla$ to get

$$u_+ = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{11.20}$$

where v_1 and v_2 are constants and

$$E^2 = \hbar^2c^2k^2 + m^2c^4 \tag{11.21}$$

This is identical to (11.2) if $\mathbf{p} = \hbar\mathbf{k}$, which is just the de Broglie relation. The components of u_- can now be obtained by substituting (11.20) into (11.16).

In the non-relativistic limit, we see, referring to (11.16), that $u_- \to 0$, and the total wavefunction is then just u_+ which is a plane wave multiplied by the two-component vector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. This has the form of a spin-half eigenvector (cf. chapter 6), the values of v_1 and v_2 determining the spin direction. We see therefore that spin is emerging in a natural way from Dirac's relativistic theory, and this point will become clearer as we proceed.

We now consider how to extend our treatment to the case where the particle is not free, but subject to an electromagnetic field represented by a scalar and a vector potential—cf. (11.3). Following the same procedure as in the free-particle case, and assuming the particle to have the electronic charge $-e$, we get, instead of (11.6),

$$i\hbar\frac{\partial}{\partial t}\psi = [c\alpha_1(\hat{P}_x + eA_x) + c\alpha_2(\hat{P}_y + eA_y) + c\alpha_3(\hat{P}_z + eA_z) + \beta mc^2 + V]\psi \tag{11.22}$$

so that (11.15) becomes

$$\left.\begin{aligned} \boldsymbol{\sigma}\cdot(\hat{\mathbf{P}} + e\mathbf{A})cu_- + mc^2u_+ + Vu_+ &= Eu_+ \\ \boldsymbol{\sigma}\cdot(\hat{\mathbf{P}} + e\mathbf{A})cu_+ - mc^2u_- - Vu_- &= Eu_- \end{aligned}\right\} \tag{11.23}$$

where $V = -e\phi$. The equivalent of (11.16) is now

$$u_- = \frac{c}{mc^2 + E - V}\big(\boldsymbol{\sigma}\cdot(\hat{\mathbf{P}} + e\mathbf{A})\big)u_+ \tag{11.24}$$

We first consider the case where the scalar potential, V, is zero. The equivalent of (11.17) is then obtained by substituting (11.24) into (11.23) to get

$$\big([\boldsymbol{\sigma}\cdot(\hat{\mathbf{P}} + e\mathbf{A})]^2\big)c^2u_+ = (E^2 - m^2c^4)u_+ \tag{11.25}$$

Expressing the left-hand side of (11.25) in Cartesian coordinates, we find that it contains terms of two types. First,

$$\sigma_x^2(\hat{P}_x + eA_x)^2c^2u_+ = (\hat{P}_x + eA_x)^2c^2u_+$$

using (11.12) and, second,

$$\begin{aligned}
&[\sigma_x\sigma_y(\hat{P}_x + eA_x)(\hat{P}_y + eA_y) + \sigma_y\sigma_x(\hat{P}_y + eA_y)(\hat{P}_x + eA_x)]c^2u_+ \\
&\quad = ie\sigma_z[A_x\hat{P}_y - \hat{P}_yA_x + \hat{P}_xA_y - A_y\hat{P}_x]c^2u_+ \\
&\quad = e\hbar\sigma_zc^2\left[Ax\frac{\partial u_+}{\partial y} - \frac{\partial}{\partial y}(A_xu_+) + \frac{\partial}{\partial x}(A_yu_+) - A_y\frac{\partial u_+}{\partial x}\right] \\
&\quad = e\hbar\sigma_z\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)c^2u_+ \\
&\quad = e\hbar\sigma_zB_zc^2u_+ \qquad (11.26)
\end{aligned}$$

where we have used (11.12) and the differential operator representation $P_x = -i\hbar\partial/\partial x$ etc. Equation (11.25) then becomes

$$\big((\hat{\mathbf{P}} + e\mathbf{A})^2 - e\hbar\boldsymbol{\sigma}\cdot\mathbf{B}\big)c^2u_+ = (E^2 - m^2c^4)u_+ \qquad (11.27)$$

The non-relativistic limit of this equation is

$$\left(\frac{1}{2m}(\hat{\mathbf{P}} + e\mathbf{A})^2 - \frac{e}{m}\mathbf{B}\cdot\hat{\mathbf{S}}\right)u_+ = \epsilon u_+ \qquad (11.28)$$

where $\hat{\mathbf{S}} \equiv \frac{1}{2}\hbar\boldsymbol{\sigma}$ represents the spin angular momentum.

The first term in (11.27) or (11.28) is what is expected for a charged particle in a magnetic field, but the second term is not. However, it has the same form as the energy of interaction between a magnetic field and a particle whose angular momentum is represented by the operator $\hat{\mathbf{S}}$ and which has a magnetic moment of magnitude $e\hbar/2m$. This is exactly what we proposed for a 'spin-half' particle in chapter 6 and we note that the free-electron g-factor comes out as equal to two automatically.

We now turn to the case of a particle moving in a scalar potential and, for simplicity, we shall assume that the vector potential is now zero. Following the same procedure as before, (11.23) and (11.24) lead to

$$\left[(\boldsymbol{\sigma}\cdot\hat{\mathbf{P}})\frac{c^2}{mc^2 + E - V}(\boldsymbol{\sigma}\cdot\hat{\mathbf{P}}) + V\right]c^2u_+ = (E^2 - m^2c^4)u_+ \qquad (11.29)$$

We define $F \equiv c^2/(mc^2 + E - V)$ and consider the first term within the square brackets in (11.29). Expressing this in Cartesian coordinates leads to two types of terms. First,

$$\sigma_x^2\hat{P}_xF\hat{P}_x = \hat{P}_xF\hat{P}_x$$

and, second,

$$\sigma_x\sigma_y\hat{P}_x F\hat{P}_y + \sigma_y\sigma_x\hat{P}_y F\hat{P}_x = i\sigma_z(\hat{P}_x F\hat{P}_y - \hat{P}_y F\hat{P}_x)$$
$$= \hbar\sigma_z\left(\frac{\partial F}{\partial x}\hat{P}_y - \frac{\partial F}{\partial y}\hat{P}_x\right)$$

using (11.12) and the differential operator representation of P_x etc.

If we now assume that the potential is spherically symmetric so that V and therefore F depend only on $r \equiv |\mathbf{r}|$, this last expression becomes

$$\hbar\sigma_z\frac{1}{r}\frac{\partial F}{\partial r}(x\hat{P}_y - y\hat{P}_x) = \hbar\sigma_z\frac{1}{r}\frac{\partial F}{\partial r}\hat{L}_z$$

where $\hat{L}_z$ represents the z component of the orbital angular momentum. Extending these results to all the Cartesian components, (11.29) becomes

$$\left[\hat{\mathbf{P}}\frac{c^2}{mc^2 + E - V}\cdot\hat{\mathbf{P}} + \frac{2}{r}\frac{\partial F}{\partial r}\mathbf{L}\cdot\mathbf{S} + V\right]c^2u_+ = (E^2 - m^2c^4)u_+ \qquad (11.30)$$

In the non-relativistic limit, we have

$$F = \frac{c^2}{2mc^2 + \epsilon - V}$$
$$= \frac{1}{2m}\left(1 + \frac{\epsilon - V}{2mc^2}\right)^{-1}$$
$$\simeq \frac{1}{2m}\left(1 + \frac{V - \epsilon}{2mc^2}\right) \qquad (11.31)$$

so that (11.30) becomes

$$\left[\frac{\hat{P}^2}{2m} + V + \frac{1}{m^2c^2}\frac{1}{r}\frac{\partial V}{\partial r}\mathbf{L}\cdot\mathbf{S} + \hat{\mathbf{P}}\cdot\frac{V - \epsilon}{4mc^2}\hat{\mathbf{P}}\right]u_+ = \epsilon u_+ \qquad (11.32)$$

This is the time-independent Schrödinger equation for a particle subject to a potential V, but with two extra terms. The third term is precisely the spin–orbit term introduced in chapter 6, while the last term is a relativistic correction of similar order, but which generally does not add significantly to the spectral structure. We note that our earlier treatment required a factor of two to be introduced into the definition of the spin magnetic moment and again in the spin–orbit expression (Thomas precession). Both of these are included in these expressions, having emerged directly from the Dirac equation. The fact that this is simply a necessary consequence of requiring consistency between relativity and quantum mechanics is one of the triumphs of the Dirac equation. We emphasize again that this does not mean that the electron is in any physical sense 'spinning'.

The whole concept of angular momentum and its conservation in a spherically-symmetric potential is a consequence of our experience with large-scale classical systems. It actually breaks down for fundamental particles where quantum effects are important, but can be re-instated if we attribute this intrinsic angular momentum to them. Of course one implication of this is that all fundamental particles should have the same value of spin and this is true for the electron, the proton and neutron and all the quarks. Other particles such as the photon and the various bosons associated with the strong and weak interactions have integer spin, but are subject to different equations.

The fully relativistic form of the Dirac equation (11.29) can be solved exactly in the case of the hydrogen atom in zero **B** field, where $V(r) = -Ze^2/(4\pi\epsilon_0 r)$ (cf. chapter 3). We shall not give the details of this here, but simply quote the result

$$E = mc^2 \left(1 + \frac{\alpha^2}{\left[n' + \sqrt{(j'^2 - \alpha^2)} \right]^2} \right)^{-1/2} \tag{11.33}$$

where $n' = n - j'$, and $j' = j + \frac{1}{2}$, n and j being the principal quantum number (cf. chapter 3) and the total angular momentum quantum number respectively ($j = l \pm \frac{1}{2}$—cf. chapter 6) and $\alpha \equiv e^2/4\pi\epsilon_0\hbar c$.

11.3 Antiparticles

The classical relativistic relation (11.2) expresses the square of the total energy E of a free particle in terms of the square of its momentum p. It follows that there is no restriction on the sign of E and the equation has a full set of solutions for E less than $-mc^2$ as well as for E greater than mc^2. Classically, these are rejected as being unphysical and this is no problem in practice as there is no mechanism for reaching the negative-energy states from the positive ones. However, in the quantum-mechanical case, negative energy states could be reached by a discontinuous quantum transition, and spontaneous transitions to states of ever-lower energy might be expected.

It is clear from (11.19) that the Dirac equation does, indeed, possess solutions with

$$E = -\sqrt{\hbar^2 c^2 k^2 + m^2 c^4}$$

To overcome the problem of transitions to negative-energy states, Dirac made the radical suggestion that these were already filled so that transitions into them would be prevented by the Pauli exclusion principle. He then considered the consequences if an electron were excited from one of these filled states with energy $-(mc^2+\epsilon)$ into a state of energy $(mc^2+\epsilon)$ under the influence of a photon of energy $2(mc^2 + \epsilon)$. The excited electron has positive energy and therefore behaves quite normally, but the filled sea of negative energy states now has a vacancy. This means that the total energy of the negative-energy states has been

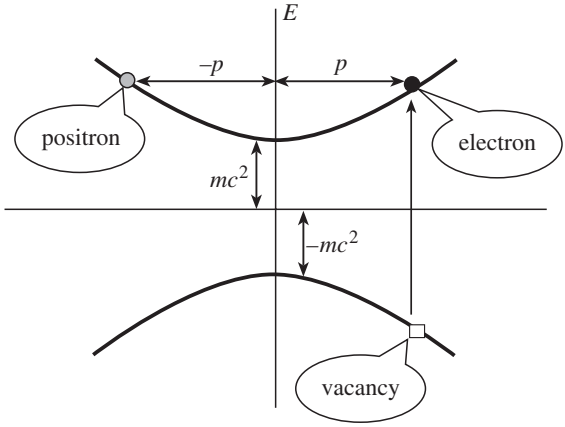


Figure 11.1. An electron can be excited from one of the filled negative energy states to create a free electron of positive energy and a vacancy. The properties of the negative-energy sea containing a vacancy are the same as those of a positron.

increased by $mc^2 + \epsilon$ and their net momentum is $-\mathbf{p}$, where $\mathbf{p}$ is the momentum of the vacancy and hence equal to that of the excited electron. Referring to figure 11.1 we consider how this momentum changes under the action of an applied electric field. All the negatively charged electrons will accelerate in a direction opposite to that of the field, taking the vacancy with them. As a result, the net momentum $-\mathbf{p}$ increases *in the same direction* as the field. It follows that the negative-energy sea plus a vacancy will behave just like a positively charged electron. This particle, which had not been observed when Dirac developed his theory, is called a 'positron', and the photon has therefore created an electron–positron pair. The experimental observation of the positron a few years after this prediction (by Anderson in 1933) was a great success for Dirac's theory. A further consequence is that all spin-half particles (protons, quarks, etc.) should have analogous 'antiparticles' and this has also been confirmed. Similar arguments to these are also used in the physics of semiconductors, where positive-charge carriers result from the excitation of electrons from otherwise full bands.

Despite its predictive success, however, there are problems with Dirac's early model. In particular, the sea of occupied negative states has no observable properties until a vacancy is created. This must mean that the infinite set of particles has no gravitational mass and no charge, despite the fact that our analysis of the expected behaviour of the vacancy assumed that all the electrons responded to an applied field as if they had charge $-e$. Another feature of Dirac's model is that there is symmetry between the positron and the electron; a theory that posited electrons as being associated with vacancies in an otherwise filled sea of positrons would make identical predictions. Later 'quantum-field' theories dispense with the idea of a filled sea and simply postulate that particle and antiparticle pairs are

excited states of the 'Dirac field' whose ground state is the vacuum. We give a brief introduction to the ideas of field theory in a later section.

11.4 Other wave equations

If, instead of following Dirac, we operate on (11.5) by

$$i\hbar\frac{\partial}{\partial t}+\sqrt{[c^2\hat{P}^2+m^2c^4]}$$

we get

$$\hbar^2\frac{\partial^2\psi}{\partial t^2}+[c^2\hat{P}^2+m^2c^4]\psi=0 \tag{11.34}$$

This is known as the *Klein–Gordon* equation, which was actually discovered before the Dirac equation. It was initially thought not to be relevant as the particle probability density associated with it is not necessarily positive. However, it was later realized that this particle density could be interpreted as a charge density, its sign depending on whether particles or antiparticles were dominant. Solutions to the Dirac equation are also clearly solutions to the Klein–Gordon equation, but the latter possesses another set of solutions with no intrinsic angular momentum. It is therefore used to describe the properties of particles with zero spin.

The relativistic equations that describe the photon are, of course, Maxwell's equations. These have to be further quantized to reveal the properties of the photon, as is discussed briefly in the next section.

11.5 Quantum field theory and the spin-statistics theorem

The spin-statistics theorem states that the wavefunctions of particles with integer spin are symmetric with respect to exchange of the labels on the particles—$\psi(1,2)=\psi(2,1)$—and obey Bose–Einstein statistics (see chapter 10). In contrast, particles with half-integer spin have antisymmetric wavefunctions—$\psi(1,2)=-\psi(2,1)$—obey Fermi–Dirac statistics, and are consequently subject to the exclusion principle. The spin-statistics theorem was shown by Pauli in 1940 to follow from some quite deep symmetry properties of relativistic quantum field theory. Many theoretical physicists over the years have believed that such a simple connection between two apparently quite separate properties of the fundamental particles should have a simple proof. A number of attempts to find such a proof have been made but, so far, none has emerged that has been generally accepted. Pauli's proof is well beyond the compass of this book, but we will give a short introduction to some of the basic ideas of field theories and explain how these give some indication of the truth of the spin-statistics theorem.

Quantum mechanics as developed up to now has taken the existence of particles such as the electron as given. In contrast, the quantum field approach builds them into the formalism of the theory itself. The starting

point is the vacuum—space and time with no matter and no radiation. This state is represented by the quantum-mechanical state vector $|0\rangle$. We then define a 'creation operator' $\hat{a}^\dagger$ which operates on $|0\rangle$ to create the state $|1\rangle$, which represents the vacuum plus one particle. Particles are removed by the corresponding 'annihilation operator', $\hat{a}$.[1]

To develop a field theory for the electron from the Dirac equation, we consider the case of a free electron with momentum $\mathbf{p} = \hbar\mathbf{k}$, so that $E = \pm E_0 \equiv \pm\sqrt{(m^2c^4 + p^2c^2)}$. A general solution of the Dirac equation is a linear combination of the wavefunctions corresponding to $\pm E_0$ and therefore has the form

$$\Psi = [au_1 \exp(-iE_0t/\hbar) + b^*u_2 \exp(iE_0t/\hbar)] \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{11.35}$$

where u_1 and u_2 are the time-independent parts of four-component vectors (cf. (11.13)) and a and b^* are constants. We note from our earlier discussion that u_1 and u_2 are orthogonal and we can also assume that they are normalized. That is,

$$\begin{aligned} u_1^\dagger u_1 &= u_2^\dagger u_2 = 1 \\ u_1^\dagger u_2 &= u_2^\dagger u_1 = 0 \end{aligned} \tag{11.36}$$

where the superscript † indicates Hermitian conjugate (cf. the discussion of matrix mechanics in chapter 6). The probabilities of finding the system in the states with energy $E = E_0$ and $E = -E_0$ are $|a|^2$ and $|b|^2$ respectively. From our earlier discussion, we expect $|a|^2$ to be the probability of the system containing an electron of momentum $\mathbf{p}$, while $|b|^2$ is the probability of a positron of momentum $-\mathbf{p}$. If we confine ourselves to low-energy states, we cause the time-dependent Schrödinger equation, and the expectation value of E is given by

$$\langle E\rangle = \left\langle \Psi^* i\hbar \frac{\partial\Psi}{\partial t}\right\rangle = E_0(a^*a - bb^*) \tag{11.37}$$

while that of the total charge is

$$\langle Q\rangle = -e\langle\Psi^*\Psi\rangle = -e(a^*a + bb^*) \tag{11.38}$$

using (11.36).

Our theory would agree with the experimental observation of electrons and positrons, if the signs of the second terms on the right-hand sides of (11.37) and (11.38) were reversed. As we shall see, this is just what quantum field theory can achieve. To move to a quantum field theory, we replace the constants a and b^* by operators $\hat{a}$ and $\hat{b}^\dagger$ and their complex conjugates by $\hat{a}^\dagger$ and $\hat{b}$ which are the Hermitian conjugates of $\hat{a}$ and $\hat{b}^\dagger$. As always, when we extend our theory into a

[1] There are separate creation and annihilation operators for every allowed value of the momentum and spin, so $\hat{a}$ and $\hat{a}^\dagger$ may be labelled accordingly. These labels are omitted in our discussion to assist clarity. Dirac notation introduced in chapter 6 is being used.

new area we do so inductively, testing the predictions of the new theory against experiment. As a^*a is a measure of the probability of the system being in a state of positive E, the operator $\hat{a}^\dagger\hat{a}$ is taken to represent the number of electrons, while the number of positrons is represented by the operator $\hat{b}^\dagger\hat{b}$. The operators $\hat{a}^\dagger$ and $\hat{b}^\dagger$ are taken to be 'creation' operators: the effect of $\hat{a}^\dagger$ is to increase the number of electrons in the state by one. Similarly, $\hat{a}$ and $\hat{b}$ are associated with the 'annihilation' of electrons and positrons respectively, reducing the number of particles in the state.

To proceed further, we have to postulate more specific properties of the creation and annihilation operators. We encountered similar operators (then called 'raising and lowering' operators) when we discussed the harmonic oscillator in chapter 4. In that case the commutation relation, $[\hat{a}^\dagger, \hat{a}] = 1$, plus the condition that all energy levels had to be positive, led to the energy spectrum $E_n = (n + \frac{1}{2})\hbar\omega$. If we assume that the same condition holds in the present case, the first term in (11.37) leads to a ladder of positive-energy levels, but the second term produces a ladder of ever decreasing negative-energy levels, which is just what we are trying to avoid.

Dirac and Jordan showed that in order to produce a field theory consistent with the Dirac equation, the operators $\hat{a}, \hat{b}, \hat{a}^\dagger$ and $\hat{b}^\dagger$ must obey 'anticommutation relations', which are similar to commutation relations, but with a positive rather than a negative sign. That is,

$$\begin{aligned} \hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a} &= \hat{b}\hat{b}^\dagger + \hat{b}^\dagger\hat{b} = 1 \\ \hat{a}\hat{a} = \hat{a}^\dagger\hat{a}^\dagger &= \hat{a}\hat{b} + \hat{b}\hat{a} = 0 \qquad \text{etc.} \end{aligned} \tag{11.39}$$

The quantities $\hat{N}_e = \hat{a}^\dagger\hat{a}$ and $\hat{N}_p = \hat{b}^\dagger\hat{b}$ represent the total number of electrons and positrons respectively. If states with other values of the momenta are included, all quantities of the type $\hat{a}_1\hat{a}_2^\dagger + \hat{a}_2^\dagger\hat{a}_1$ etc. are equal to zero.

We first establish the creation and annihilation properties of $\hat{a}^\dagger$ and $\hat{a}$, noting that identical arguments can be applied to $\hat{b}^\dagger$ and $\hat{b}$. From (11.5) we have

$$\begin{aligned} \hat{a}\hat{a}|\rangle &= 0 \\ \hat{a}^\dagger\hat{a}^\dagger|\rangle &= 0 \end{aligned} \tag{11.40}$$

using Dirac notation (cf. chapter 6). There are only two states, $|0\rangle$ and $|1\rangle$, that satisfy these relations and therefore obey the conditions

$$\begin{aligned} \hat{a}|0\rangle &= 0 \qquad \hat{a}|1\rangle = |0\rangle \\ \hat{a}^\dagger|1\rangle &= 0 \qquad \hat{a}^\dagger|0\rangle = |1\rangle \end{aligned} \tag{11.41}$$

which directly lead to

$$\begin{aligned} \hat{N}_e|0\rangle &= \hat{a}^\dagger\hat{a}|0\rangle = 0 \\ \hat{N}_e|1\rangle &= \hat{a}^\dagger\hat{a}|1\rangle = |1\rangle \end{aligned} \tag{11.42}$$

so the eigenvalues of $\hat{N}_e$ are $n_e = 0$ (corresponding to an empty state) and $n_e = 1$ (a filled state). Remembering that we are currently considering states with particular values of $\mathbf{k}$ and $\boldsymbol{\sigma}$, there are only two allowed states of the system: one with no electrons and one with a single electron, which is just the Pauli exclusion principle. Similar arguments using $\hat{b}$ and $\hat{b}^+$ produce corresponding results for the position states. We can also see how field quantization leads to the antisymmetric properties of the wavefunction. Let $|1, 2\rangle$ be the ket representing the state of two electrons, labelled 1 and 2. We can generate this by the operation of the creation operators on the vacuum state $|0\rangle$:

$$|1, 2\rangle = a_1^\dagger a_2^\dagger |0\rangle$$

Hence

$$|2, 1\rangle = a_2^\dagger a_1^\dagger |0\rangle = -|1, 2\rangle \tag{11.43}$$

where the last step uses (11.39).

Returning now to the properties of the state (11.35), the total energy (11.37) becomes

$$\langle E\rangle = E_0\langle \hat{a}^\dagger\hat{a} - \hat{b}\hat{b}^\dagger\rangle = E_0\langle \hat{a}^\dagger\hat{a} + \hat{b}^\dagger\hat{b} - 1\rangle = E_0(n_e + n_p) \tag{11.44}$$

where we have assumed that the system is in an eigenstate of $\hat{N}_e$ and $\hat{N}_p$ with eigenvalues n_e and n_p respectively. We can similarly express the total electric charge (cf. (11.38)) as

$$\langle Q\rangle = -e\langle \hat{a}^\dagger\hat{a} + \hat{b}\hat{b}^\dagger\rangle = -e(n_e - n_p) \tag{11.45}$$

We see that $\langle E\rangle$ is just the sum of the energies of the electrons and positrons, which are both positive. Moreover, $\langle Q\rangle$ is just the expected net charge. It should be noted that we have omitted a term $-E_0$ in (11.44) and $-e$ in (11.45) in order to ensure that the energy and charge of the vacuum state are both zero. Such 'zero-point' terms are infinite when totalled over all the energy states: the subtraction of such infinities in order to obtain a physical result is a common feature of more advanced aspects of quantum-field theory.

Summarizing, we have shown that a field theory which associates positive energy with the existence of both electrons and positrons requires the creation and annihilation operators to obey anticommutation relations. This, in turn, ensures that the occupation numbers of the states are either zero or one, so that the Pauli exclusion principle is obeyed.

As stated earlier, this argument leading to the Pauli exclusion principle does not constitute a complete proof. One reason is that it is much harder in the more general case when the effects of interactions are included. It also says nothing about the properties of particles with integer spin where Bose–Einstein statistics are expected to apply. However, if field quantization is applied to the Klein–Gordon equation in the spin-zero case, it is found that positive energies for both particles and antiparticles are obtained assuming that the creation and annihilation

operators obey commutation rather than anticommutation relations. These permit multiple occupation of the states and hence Bose–Einstein statistics. To complete the proof, we also have to show that fermion-type solutions to the Klein–Gordon equation, that are not also solutions to the Dirac equation, are not allowed and this is considerably harder. A further problem is that the Dirac equation relates only to spin-half and the Klein–Gordon equation only to spin-zero particles, while the spin-statistics theorem refers generally to even and odd numbers of half-integers. However, particles with spin greater than one-half may be properly described as tightly bound composites of spin-half particles, which would then obey the spin-statistics theorem.

A fundamental feature of all these approaches to the spin-statistics theorem is that they rely on the relativistic regime to predict a result that applies to systems where relativistic effects are otherwise negligible. This has prompted quite different approaches to the problem, notably a suggestion by Berry and Robbins in 1997 that the antisymmetry of the two-fermion wavefunction may be associated with a geometrical phase factor resulting from the topology of the system.

As was pointed out in chapter 1, quantum mechanics began with the idea of the quantum of electromagnetic radiation (or photon), but this is properly included in quantum mechanics only after Maxwell's equations for the electromagnetic field are quantized in a similar way. The resulting photons have the energy and momentum given by the Planck and de Broglie relations and are found to have total-spin quantum number $j = 1$. The states with $m_j = \pm 1$ correspond to right- and left-circularly polarized plane waves; the state with $m_j = 0$ would correspond to a longitudinally polarized electromagnetic wave and is forbidden. The interaction between the fields representing matter and radiation can also be built into quantum field theory and new results have been predicted that have been experimentally confirmed.

As we indicated earlier, the essential advantage of quantum field theory is that it includes the existence of particles in the formalism as quanta of the field, whereas previously we assumed the existence of, say, an electron whose quantum properties were described by the Schrödinger equation. The concept of the quantized field also enables an alternative and, perhaps, deeper understanding of the concept of indistinguishability. Rather than saying that 'particles 1 and 2 are identical' we can simply say that the field contains two excitations which we do not attempt to label. An analogy is sometimes drawn between having two identical pound coins and two pounds in a bank account. In the latter case the two units have no individual identity and field theory provides a similar conceptual basis for the description of a field containing two quanta.

Problems

11.1 Show by substitution that the matrices given in (11.9) have the properties set out in (11.8).

11.2 Show that if $E \gg mc^2$, the energy eigenstates of a spin-half particle have definite helicity—i.e.

they are in eigenstates of the spin component parallel to the momentum direction.

11.3 Show that in the limit where α is small, the relativistic energy levels (11.33) are equal to those obtained for the hydrogen atom in chapter 3 (3.65).

11.4 Obtain an expression for the wavefunction of a free particle with negative energy by using the first of (11.15) to express u_+ in terms of u_- and hence obtain a version of (11.20). Hence show that the wavefunctions corresponding to the same values of $\mathbf{k}$ and v but with oppositely signed E are orthogonal.

11.5 Explain why the field operator representing the scattering of particles from states labelled 1 and 2 into states labelled 3 and 4 is

$$c_3^\dagger c_1 + c_4^\dagger c_2$$

Chapter 12

Quantum information

This chapter aims to provide a first introduction to some of the new applications of the ideas of quantum mechanics that have been developed during the last twenty years or so of the twentieth century. The topics chosen reflect our increased understanding of a number of concepts and their application to new phenomena that have been predicted and sometimes observed. They all rely on the interplay between the two types of time dependence implicit in the fundamental postulates set out in chapter 3:

Postulate 5 Between measurements, the development of the wavefunction with time is governed by the time-dependent Schrödinger equation.

Postulate 2 Immediately after a measurement, the wavefunction of the system will be identical to the eigenfunction corresponding to the eigenvalue obtained as a result of the measurement.

There is generally a clear distinction between 'unitary evolution' under the influence of the time-dependent Schrödinger and the 'collapse' associated with a measurement. In unitary evolution (so called because in a matrix representation, the initial and final states are connected by a unitary matrix) the final state of the system is completely determined by the Hamiltonian operator and the initial state. It follows that unitary evolution is reversible as the initial state can be generated from the final state by the same time-dependent Schrödinger equation with the sign of the time coordinate reversed and ψ replaced by ψ^*. In contrast, the result of a measurement on quantum mechanics is generally unpredictable; the system 'collapses' at random into one of a set of possible outcomes. Thus collapse is an irreversible change that occurs in one time direction only. Although in practice we easily know when to apply which form of time dependence, it is very difficult to set objective criteria for this. We shall discuss the consequent 'quantum-mechanical measurement problem' in more detail in chapter 13.

All the examples to be discussed in this chapter refer to the behaviour of two-state systems. A prime example of these is the spin-half particle, whose

component of angular momentum relative to, say, the x axis can have the values $\pm\frac{1}{2}\hbar$. Moreover, if we choose a different axis of quantization, the new eigenfunctions can be expressed as linear combinations of the old. For example, if α_z and β_z represent positive and negative z components respectively while α_x and β_x represent the corresponding quantities measured with respect to the x direction,

$$\begin{aligned}\alpha_x &= 2^{-1/2}(\alpha_z + \beta_z)\\ \beta_x &= 2^{-1/2}(\alpha_z - \beta_z)\end{aligned} \tag{12.1}$$

cf. chapter 6 (6.23). Another example of a two-state system is a plane-polarized photon. Plane polarization is a familiar concept in classical optics where it refers to the plane in which the E vector of the electromagnetic wave vibrates. When experiments are done on weak light, it is found that this property can also be applied to individual photons. The two states of polarization follow very similar rules to the spin directions in the spin-half case. For example, the states of being polarized at either plus or minus 45° to the horizontal axis are related to the horizontally and vertically polarized states by an equation of the same form as (12.1).

Atoms can also be used as two-state systems despite the fact that they generally have an infinite number of bound states! This is achieved by ensuring that they are isolated from all external influences other than radiation whose frequency (typically in the microwave region) matches the energy difference between two of the states. To minimize the frequency of unwanted collapses due to spontaneous transitions between the states, systems are chosen where the energy difference is very small (cf. chapter 8).

12.1 Quantum cryptography

Cryptography is the science of the exchanging of messages in a coded form that makes them indecipherable to anyone else. We shall not attempt anything like a full survey of this immense subject, but will concentrate on how the particular properties of quantum systems can be used in this area. To transmit any message at all, it must be coded in some form. The pages in this book do this using the alphabet and other symbols, as well as relying on the reader's and writer's knowledge of the English language and mathematics. When the present edition was written using a computer with a word-processing package, the letters, figures and symbols were further encoded by the computer into a set of binary numbers represented by two symbols '1' and '0'. In this way any message at all can be represented by a sequence of 1s and 0s, and any system that is capable of existing in two distinct states can be used to encode such a representation of a message. This procedure is now universally used in electronic transmission where the 1s and 0s are often represented by voltage pulses of different sizes.

The essential property of an encrypted message is that the information in it should be understandable only to the sender and the receiver, and be meaningless to a third party. This is achieved by processing the message using some mathematical procedure known as a 'cypher', which can be decoded only if the reader is in possession of a 'key'. Early 'keys' were in the form of code books, but modern cryptography uses mathematical procedures that depend on knowing a comparatively small key, typically a few thousand binary bits in length.

For example, a message, M, could be coded as E by the algorithm

$$E = M^s \mod c \tag{12.2}$$

where the right-hand side means that the number M is raised to the power s and the answer expressed as a number of base c. If c is a product of two prime numbers (p and q), it can be shown that the message can be decoded by the algorithm

$$M = E^t \mod c \tag{12.3}$$

where t is a simple function of p and q. Provided p and q are kept secret, all the other quantities can be exchanged publicly. If c is large enough, breaking the code by searching for the prime factors , p and q, would take a fast conventional computer an impossibly long time (more than 10^6 years if c is 1000 binary digits in length). As we shall see later in this chapter, this factorization problem is one where quantum computers could, in theory, make a dramatic contribution, but it may well be 10^6 years before this technology is available!

The present section describes a method whereby a sender and receiver can both acquire knowledge of a number, while being sure that no eavesdropper knows it. In this way they can both follow the same coding process, even if they do not know what this is going to be before the exchange takes place. This key distribution process is where quantum mechanics comes in.

A sender (conventionally known as 'Alice') could transmit a message to a receiver ('Bob') in the form of a stream of photons with an appropriate sequence of polarization directions, say horizontal (H) for 0 and vertical (V) for 1. However, there is no obvious reason why the message should not be intercepted by an eavesdropper ('Eve') and re-transmitted to Bob without either of them knowing that this had happened. We can avoid this by exploiting the specifically quantum properties of the particles.

The first protocol aimed at this end was suggested in the 1980s. The central idea is that Alice, instead of sending just H or V photons, sends a sequence where some are H/V, but others are $\pm 45°$. Both H and $45°$ are to be interpreted as 0, while V and $-45°$ represent 1. Which orientations Alice uses are decided at random, but she keeps a record of them. When Bob detects the photon he also varies the orientation of his apparatus between H/V and $\pm 45°$ at random and records which apparatus he uses as well as the results he obtains. Alice and Bob then openly tell each other which settings they used in each case and discard those where these were not the same—on average half the total. The remainder consists

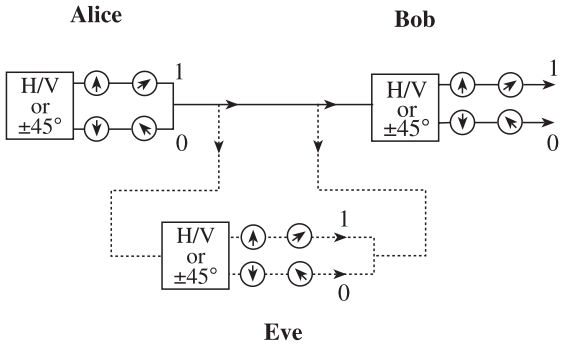


Figure 12.1. Alice sends a series of photons to Bob, which are polarized either H/V or ±45° at random. Bob makes a similar set of randomly selected measurements on the photons he receives. They then exchange information on the measurement orientations they have chosen and use the results for the subset where these are the same, to define the key to be used. In the absence of Eve's intervention, Alice and Bob have the same key, but this no longer holds if Eve makes an intermediate measurement.

of a set of numbers whose values they both know and which they can then use to decide which procedure they will follow to encode and decode messages sent openly between them.

We now consider the effect of the possible intervention of Eve. She may intercept the photons sent by Alice, but she has no means of knowing the orientation Alice used when sending them. The simplest thing she can do is to randomly guess which orientation might have been used for a particular photon, set her apparatus in this orientation, record the result and send it on to Bob. On average, she will guess right one-half of the time, but there is only a 50% probability that this guess will be the same as Bob's. The upshot is that one-half of the photons that Bob later believes are reliable, have been corrupted by passing through Eve's apparatus set in the 'wrong' orientation. About one-half of these (i.e. one-quarter of the final sequence used as the key) will decode as bits that are the opposite of those sent by Alice. Alice and Bob therefore no longer have a set of numbers in common and soon find they cannot decode each other's messages. They have therefore detected the presence of Eve and can take appropriate action. An illustration of how this process works is shown in figure 12.1 and table 12.1.

This protocol succeeds because of the collapse resulting from a quantum measurement. A necessary concomitant of this is the system's vulnerability to random processes (noise). Even in the absence of Eve, Alice and Bob will successfully end up with a common key only if their apparatuses are accurately aligned when they believe they are, and if the polarization state has not been changed (e.g. rotated slightly) during the transmission. Moreover, even though Eve cannot accurately read the message and her attempt to do so can be detected,

Table 12.1. The properties of a typical set of 12 photons analysed by the apparatus shown in figure 12.1. The asterisks in the final column mark cases where Eve's intervention has resulted in Bob having the wrong result. (AS, Alice's polarizer setting; PA, polarization of photons sent by Alice; AM, Alice's message; BS, Bob's analyser setting; BR, Bob's result; AR, accept/reject; K, common key; ES, Eve's polarizer setting; PE, polarization of photons sent by Eve.)

Alice			Bob (Eve off)				Bob (Eve on)			
AS	PA	AM	BS	BR	AR	K	ES	PE	BR	K
H/V	H	0	H/V	H	y	0	H/V	H	H	0
H/V	V	1	H/V	V	y	1	±45°	–	–	0
±45°	+	0	H/V	H	n		H/V	H	H	
H/V	H	0	±45°	–	n		H/V	H	–	
±45°	+	0	±45°	+	y	0	H/V	H	–	1*
±45°	–	1	H/V	V	n		±45°	–	V	
H/V	V	1	±45°	–	n		±45°	–	–	
±45°	–	1	H/V	V	n		H/V	H	H	
±45°	+	0	±45°	+	y	1	H/V	V	+	0*
±45°	+	0	H/V	H	n		±45°	+	H	
H/V	H	0	H/V	H	y	0	H/V	H	H	0
H/V	V	1	H/V	V	y	1	±45°	V	–	1*

her presence apparently makes the channel of communication useless. There are strategies for combating this by further public exchange of information, as a result of which Alice and Bob can end up with a shared key considerably shorter than the original, but with a very low probability of Eve also knowing it.

Unlike the cases to be discussed in the rest of this chapter, quantum cryptography is quite easy to carry out experimentally. Secure key interchange has been demonstrated by sending polarized photons over distances up to about 20 km of standard optical communication fibre.

12.2 Entanglement

The rest of the examples we discuss involve two or more particles in what is called an 'entangled state'. In quantum mechanics, the word 'entanglement' refers to a quantum state of two or more particles in which the probabilities of the outcome of measurements on one of them depend on the state of the other, even though there is no interaction between them.

As an example of entanglement, consider two spin-half particles in a state where their spins are opposite, but we have no other information about them. The

spin part of their wavefunction has the form

$$\psi(1,2) = 2^{-1/2}[\alpha_z(1)\beta_z(2) - \beta_z(1)\alpha_z(2)] \tag{12.4}$$

where $\alpha_z(i)$ and $\beta_z(i)$ represent particle i with a positive and negative spin component respectively relative to the z axis. However, the only information this function contains is that the spins are opposite—it does not tell us the absolute direction of either spin. This is because $\psi(1,2)$ is independent of the direction of the axis of quantization. To see this, we apply the transformation (12.1) to express $\psi(1,2)$ in terms of $\alpha_x(i)$ and $\beta_x(i)$ and yet:

$$\psi(1,2) = -2^{-1/2}[\alpha_x(1)\beta_x(2) - \beta_x(1)\alpha_x(2)] \tag{12.5}$$

which is the same as (12.4) apart from an irrelevant change of sign. We therefore drop the suffixes on α and β from now on when we describe states entangled in this way.

We now consider the effect of performing a measurement of the z component of either of the spins when the system is in the state $\psi(1,2)$. Clearly we expect to get a positive or a negative result at random with equal probability. But suppose we measure the spin of particle 2 *after* we have measured the spin of particle 1 and obtained (say) a positive result. As a result of this first measurement, the system must have collapsed into an eigenstate of the spin-component operator with positive eigenvalue. It therefore has the form

$$\psi(1,2) = \alpha_z(1)\beta_z(2) \tag{12.6}$$

and we now do know that the particles are in eigenstates of $\hat{S}_z$, because the state is now disentangled. A measurement of the z component of the spin of the second particle can now only yield a negative result. It follows that the probabilities of obtaining particular values of the spin of one particle depend on what measurements have been previously carried out on the other. It should be noted that this result is independent of other properties of the particles, in particular their position: particles can be entangled even when a long distance apart. This apparent influence of the operations carried out on one particle on the properties of a distant particle is known as 'non-locality' and we discuss its implications for our understanding of the conceptual basis of quantum mechanics in chapter 13.

12.3 Teleportation

The word 'teleportation' entered the English language via the science-fiction series *Star Trek*. In this fantasy, a person or object could be transported to a distant destination using a transmitter that measured all the properties of the object and sent them (presumably by some radio link) to a receiver that re-assembled the information to re-create the object. In the process the original object was

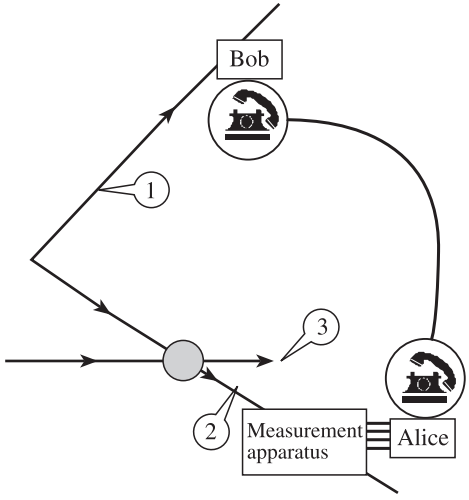


Figure 12.2. Quantum teleportation. Particles 1 and 2 are prepared in an entangled state. Alice allows particle 2 to interact with particle 3, then measures a property of particle 2 and uses a classical communication channel to tell Bob which of the four possible results she obtained. Bob can then transform particle 1 into a state identical to the initial state of particle 3.

destroyed (de-materialized?). For a long time it was believed that such a process was not only impossibly difficult in practice, but was forbidden in principle by quantum mechanics. This is because the uncertainty principle ensures that we can never determine all the properties of the state of a quantum system, and hence all the information that would be needed to reproduce it. This principle is sometimes called the 'no-cloning' theorem, which states that it is impossible to produce a particle with exactly the same quantum properties as another without altering the state of the initial particle. More recently, however, it has been realized that it is possible to use entanglement to copy the properties of one particle onto another at a distant destination, without making a record of them in the process, although inevitably changing the quantum state of the original particle. This process has become known as 'quantum teleportation'.

We consider the simplest possible example of a spin-half particle (which we shall call particle 3) in a state given by

$$\psi(3) = A\alpha(3) + B\beta(3) \tag{12.7}$$

where the values of A and B are unknown. An experimentalist (Alice of course) wants to transfer these values to a different particle, which she has sent to another experimenter (Bob). We assume that Alice also has available a pair of previously entangled particles (1 and 2) whose quantum state is given by (12.4). The total

wavefunction of all three particles is therefore

$$\Psi(1,2,3) = 2^{-1/2}[A\alpha(3) + B\beta(3)][\alpha(1)\beta(2) - \beta(1)\alpha(2)] \tag{12.8}$$

This is algebraically identical to the expression

$$\begin{aligned}\Psi(1,2,3) = &-\tfrac{1}{2}[A\alpha(1) + B\beta(1)]\psi_1(2,3) + \tfrac{1}{2}[A\alpha(1) - B\beta(1)]\psi_2(2,3)\\ &- \tfrac{1}{2}[A\beta(1) + B\alpha(1)]\psi_3(2,3) - \tfrac{1}{2}[A\beta(1) - B\alpha(1)]\psi_4(2,3)\end{aligned} \tag{12.9}$$

where

$$\begin{aligned}\psi_1(2,3) &= 2^{-1/2}[\alpha(2)\beta(3) - \beta(2)\alpha(3)]\\ \psi_2(2,3) &= 2^{-1/2}[\alpha(2)\beta(3) + \beta(2)\alpha(3)]\\ \psi_3(2,3) &= 2^{-1/2}[\alpha(2)\alpha(3) - \beta(2)\beta(3)]\\ \psi_4(2,3) &= 2^{-1/2}[\alpha(2)\alpha(3) + \beta(2)\beta(3)]\end{aligned} \tag{12.10}$$

as can be checked by expanding the right-hand sides of (12.8) and (12.9).[1]

Alice sends particle 1 to Bob and then performs a measurement on particles 2 and 3. The essential feature of this measurement is that it is represented by an operator whose eigenfunctions are the four ψ_is, so that the whole system, including particle 3, collapses into one of the terms in (12.10). Alice communicates the result of this to Bob, using a *classical* channel of communication—cf. figure 12.2. Thus, if Alice obtains the result corresponding to ψ_1, Bob's particle (1) has the same spin function as was originally possessed by particle 3, which of course was Alice's original intention. If she obtains one of the other results and lets Bob know, then Bob can transform the state function into that required simply by rotating the spin using a magnetic field B. For example, to generate the second square-bracketed expression from the first, we use (8.13) in chapter 8 to get

$$A(t) = A(0)\exp(-\tfrac{1}{2}i\omega_p t) \text{ and } B(t) = A(0)\exp(\tfrac{1}{2}i\omega_p t) \tag{12.11}$$

where $\omega_p = eB/m$. We get the required result (apart from an irrelevant phase factor) if $t = \pi m/eB$.

We note that a classical communication channel by which Alice communicates the result of her measurement to Bob is an essential part of the experiment. This is an example of a general principle that forbids the transmission of information using entanglement alone. Otherwise we could transmit information instantaneously in breach of the principles of special relativity. We shall return to this point in chapter 13.

[1] This set of states (known as 'Bell states' after John Bell—see chapter 13) can be shown to have the maximum possible entanglement for two spin-half particles.

The principles of quantum teleportation have been experimentally demonstrated using atomic beams and also in the case of particular photon states. However, before any significantly complex object could be teleported, it would have to be coupled in some way to a system consisting of as many entangled particles as its relevant properties. The practicalities of this are certainly well beyond our present technology.

12.4 Quantum computing

A conventional computer essentially manipulates binary bits according to a set of rules, the operation performed on one bit often depending on the state of one or more of the others. In a 'quantum computer' the binary bits are two-state quantum systems known as 'qubits' and the operations proceed by unitary evolution (i.e. the states of the qubits evolve according to the time-dependent Schrödinger equation) until we make measurements on them. It has been shown that there is no obstacle *in principle* to the construction of a quantum computer that would operate on qubits in the same way as a conventional computer processes conventional bits. However, the emphasis is on the phrase 'in principle'. As we shall see, quantum computers imply the entanglement of large numbers of qubits; such entangled states are extremely sensitive to noise and decoherence, so the practical obstacles to constructing a useful device are immense and would imply a completely new technology.

Of course there is no reason why anyone would go to such trouble to construct something that could do no more than a conventional computer. The great interest in quantum computers is that, although they would carry out most computing tasks no more efficiently than conventional machines, there are some processes that would be performed at an immensely faster speed. The example most often quoted is the factorization of a large number into its prime-number components. As mentioned earlier, this operation plays an important role in some cryptographic protocols. For a number consisting of N decimal digits, the best known conventional method requires a number of computing steps that is proportional to $\exp(2L^{1/3}\ln L^{2/3})$, where $L = \ln N$. In contrast the number of operations required by a quantum computer has been shown to be about $300L^3$. For L equal to 10, the classical algorithm is faster, but the quantum computer takes over for larger numbers: for $L = 200$, the calculation would take 10^9 years classically, but could be performed by a quantum computer in about 8 hours.

As a first step to understanding the principles of quantum computation, we consider a particular simple operation, known as a NOT gate. This has the property of reversing the state of a bit: if we input 0 we get out 1, while if we input 1 we get out 0. It is very easy to devise a unitary operation that will carry out this procedure on a qubit, represented by a spin-half atom. We simply pass the spin through a magnetic field, adjusted so that it rotates the spin through 180° in the process. The important thing to note is that the same operation is performed if the

initial state is a linear superposition of the states. Thus if 0 and 1 are represented by α and β, and we pass the spins through the same field as defined in the previous section (cf. (12.11)), we get not only

$$\text{and} \qquad \begin{aligned} \alpha = 2^{-1/2}\begin{bmatrix}1\\1\end{bmatrix} &\to -i2^{-1/2}\begin{bmatrix}1\\-1\end{bmatrix} = -i\beta \\ \beta = 2^{-1/2}\begin{bmatrix}1\\-1\end{bmatrix} &\to -i2^{-1/2}\begin{bmatrix}1\\1\end{bmatrix} = -i\alpha \end{aligned} \tag{12.12}$$

but also

$$A\alpha + B\beta \to -i(A\beta + A\alpha) \tag{12.13}$$

where A and B are any constants. This single NOT gate can be considered as a very simple computer program, if we imagine that we do not know in advance what the result of passing data (in this case a single bit) through it will be. If this were a conventional device, we could fully determine its properties only by running it twice: once with each of its possible inputs (0 or 1). In the quantum case, however, if we run the program once only using the left-hand side of (12.13) as input, then the wavefunction of the output—i.e. the right-hand side of (12.13)—is a linear combination of both outcomes. Thus, both calculations have been performed in one step. Before getting carried away by this apparently miraculous result we should realize that, in order to access this information, we have to make a measurement, which inevitably involves collapse. This means that we can determine only one component of the spin, so there is apparently no practical advantage in performing the calculation in this way. For this reason, it is generally true that quantum computation has little or no advantage over conventional methods. However, there are exceptions to this rule when a subtle interplay of unitary evolution and collapse can be employed to perform particular calculations with hugely improved efficiency. One of these is the determination of the period, r, of a periodic function $f(x)$ of a variable x. For reasons we shall not go into, this is a key part of the algorithm to factorize a product of two prime numbers. We shall give an outline of the principles of how r might be obtained, provided that the correct unitary operations could be performed and the appropriate measurements made.

We assume that we know that $N/2 < r < N$ for some N and define $n = 2\ln(N)$. If we have $2n$ qubits, we could use n of them (known as the 'x register') to code every number from 0 to $w-1$, where $w = 2^n$ and use another n to code the corresponding values of $f(x)$ (the 'y register'). Table 12.2 illustrates this for the case where $n = 3$ and $f(x) = 7\cos(2\pi x/3))$. However, what we actually do is first to set all the qubits to 0 (i.e. in a quantum state α) and then subject the x register to a coherent operation that leaves it in the state

$$\psi(1, 2, \ldots, n) = \frac{1}{\sqrt{w}} \sum_{i=0}^{w-1} \eta_i(1)\eta_i(2)\ldots\eta_i(n) \tag{12.14}$$

Table 12.2. The x register consists of qubits 1–3 representing values of x from 0–7, while the y register consists of qubits 4–6 which can code the corresponding values of $f(x)$. When a linear superposition of the states in column 2 is used to calculate f, the outcome is a linear combination of the products of the states in columns 2 and 3. A measurement of the states of the column-3 particles causes a collapse into either the state $\beta\beta\beta$ or $\alpha\beta\alpha$. If the first result is obtained, the first three qubits will be in a linear combination of the states listed in column 4; these clearly possess the periodicity of $f(x)$ and the same is true if the alternative result is obtained. A measurement of the new state of the x register yields a result, from which the period can usually be deduced.

$x \equiv i$	x register $j = 1, 2, 3$	y register $j = 4, 5, 6$	x register after 1st measurement $j = 1, 2, 3$	x register after F.T. $j = 1, 2, 3$	x register after 2nd measurement $j = 1, 2, 3$
0	$\alpha\alpha\alpha$	$\beta\beta\beta$	$\alpha\alpha\alpha$	$\alpha\alpha\beta$	
1	$\alpha\alpha\beta$	$\alpha\beta\alpha$			
2	$\alpha\beta\alpha$	$\alpha\beta\alpha$			
3	$\alpha\beta\beta$	$\beta\beta\beta$	$\alpha\beta\beta$	$\alpha\alpha\beta$	
4	$\beta\alpha\alpha$	$\alpha\beta\alpha$			
5	$\beta\alpha\beta$	$\alpha\beta\alpha$			
6	$\beta\beta\alpha$	$\beta\beta\beta$	$\beta\beta\alpha$	$\alpha\alpha\beta$	$\alpha\alpha\beta$
7	$\beta\beta\beta$	$\alpha\beta\alpha$			

where $\eta_i(j)$ is the state (α or β) of the jth qubit of the number i. For example, in the case where $n = 3$, $\psi(1, 2, 3)$ is proportional to the sum of the terms in the second column of table 12.2. We note that all the qubits in the x register are now in an entangled state. The next step is to subject the $2n$ qubits to another coherent operation, the outcome of which is to place them all in an entangled state. To do this, we assume that we have a computer program that can calculate $f(x)$ for any given value of x by a sequence of unitary operations. If we were to input a number such as that represented by *one* of the states listed in the second column of table 12.2, the state of these n qubits would be unchanged, but the state of the second set would evolve to represent the value of $f(x)$ as in column 3 of table 12.2. What we actually do is input the linear superposition (12.14) of x-register states. By standard quantum mechanics, the wavefunction of all $2n$ qubits evolves into

$$\psi(1, 2, \ldots, 2n) = \frac{1}{\sqrt{w}} \sum_{i=0}^{w-1} \eta_i(1)\eta_i(2)\ldots\eta_i(n)\eta_i(n+1)\ldots\eta_i(2n) \quad (12.15)$$

For $n = 3$, this is just the sum of the products of the corresponding terms in columns 2 and 3 in table 12.2.

It is important to note that this last operation does *not* involve calculating

each term separately as it would classically. By performing a single calculation, we have created a linear combination of the results corresponding to all w values of x. Of course this does not mean that we can recover and use all this information, because we cannot measure all the properties of a quantum state. Consider, however, what happens if we measure the state of the n qubits in the y register. By the measurement postulate, this collapses the whole wavefunction onto the subset of states corresponding to the eigenvalue obtained. However, because of the periodic property of $f(x)$, this will not be a single state, but a linear combination of all the degenerate states in the x register that correspond with the obtained eigenvalue of the y register. It follows that the numbers represented by the surviving x-register states have acquired the periodicity of $f(x)$. We now subject the x register to a unitary Fourier transformation as a result of which the only non-zero contributions to the superposition are those with $i = mw/r$, where r is the period we are trying to find and m is an unknown integer. We now measure the state of the x register and obtain a value for i. We perform a classical calculation to cancel down i/w to get a result which will equal r if m and r have no common factors. We can then check conventionally whether f does have period r and, if not, we can run the process again. If the numbers are large, the chance of m and r having a common factor is small and the likely number of repetitions needed to obtain the correct answer can be shown to be much less than $\ln r$.

This example illustrates the enormous potential power of the quantum computer, the essential feature being the almost miraculous way in which a superposition of the results of n calculations can be generated by a single operation. The measurement process limits the amount of this information we can retrieve, but in an appropriate case such as that described here, incredible results can be achieved. Before getting carried away, however, we should remember the immense obstacles that lie in the way of any practical application of these ideas. Although entangled states of two particles can be generated almost routinely, extending this to more than a few is difficult, and several hundred or more really seems impossible. However, there are many examples in the past of the seemingly impossible being achieved and we should beware of underestimating human technological capacity. Whether more 'in principle' objections may arise if and when we understand the measurement process better will be speculated on in the next chapter.

Problems

12.1 Construct a variant of table 12.1 in which Eve makes different guesses about the settings of Alice's apparatus.

12.2 Confirm that equation (12.5) is the same as (12.4).

12.3 Construct the equivalent of table 12.2 in the case of a quantum computation based on four two-state particles.

Chapter 13

The conceptual problems of quantum mechanics

The previous chapters contain many examples of the successful application of quantum mechanics to the solution of real physical problems. These represent only a small sample of the wide range of experimental results, spanning just about all areas of physics and chemistry and beyond, which have been successfully explained by quantum theory. So far at least, no quantum-mechanical prediction has been experimentally falsified. Despite these successes, however, many scientists have considered the basic conceptual framework of the subject to be unsatisfactory, and repeated attempts have been made to reinterpret quantum mechanics, or even replace it with a different theory whose philosophical and conceptual basis could be considered more acceptable. In the present chapter we shall explain the reasons for this dissatisfaction and outline some of the reinterpretations and alternative approaches that have been devised. Inevitably many of the questions that arise in this area are matters of opinion rather than fact and for this reason some physicists consider that they belong more properly to the realms of philosophy than of physics. However, the conceptual basis of quantum mechanics is so fundamental to our whole understanding of the nature of the physical universe that it should surely be important for physicists to understand the nature of the problems involved, if nothing else.[1,2]

13.1 The conceptual problems

In this section we outline some of the main areas of difficulty with a view to more detailed discussion later in the chapter.

[1] A more extended and less technical discussion of some of these ideas can be found in my book *Quantum Physics: Illusion or Reality* (Cambridge University Press, Canto Edition 1994).

[2] Some of the ideas in this chapter are also discussed in chapter 12, but the present account is self-contained.

Determinism

One of the ways in which quantum mechanics differs from classical physics is that the latter is a deterministic theory, which means that the laws are framed in such a way that each event can be seen as a necessary consequence of the theory and the preceding state of affairs. To take a simple example, if we release a classical object (e.g. a ball) in a vacuum, it will certainly fall to the floor and we can calculate the time it will take to do so to a very high accuracy—given the initial position of the object relative to the floor and the acceleration of gravity. In contrast, a given state of a quantum system generally allows us to predict only the relative probabilities of different outcomes. For example, if a spin-half particle, known to be in an eigenstate of $\hat{S}_x$, enters a Stern–Gerlach apparatus oriented to measure $\hat{S}_z$, we know that a positive and a negative result are equally likely, but which actually occurs is unpredictable.

It is important to realize that, although quantum physics is generally applied to the microscopic world of atoms and the like, indeterminism can also affect macroscopic events. Thus, if we have detectors in each channel of the Stern–Gerlach experiment described earlier, the thing that is uncertain is which of these will 'click' and record the passage of an atom. This counter click is a perfect macroscopic event and could be used to trigger any other everyday action. For example, we could decide that if one counter clicked, we would go to work, while if the other did so we would go back to bed. The possibly earth-shattering consequences of this decision are then completely indeterminate and cannot be predicted in advance by any observation we might make on the system before the experiment is performed.

The indeterminism of quantum mechanics can be traced back to the measurement postulate (postulate 4.4 in chapter 4) where the procedure for predicting the relative probabilities of different outcomes of a measurement is defined. It is important to note that it is here, in the measurement, that the uncertainty arises. In contrast, the 'unitary' evolution of the wavefunction between measurements is perfectly deterministic. We mean by this that if we know the wavefunction at any time, we can in principle, and in the absence of measurement, use the time-dependent Schrödinger equation to calculate its form at any future time (see chapter 8 and the discussion of the measurement problem later in this chapter). Referring again to the Stern–Gerlach example, the wavefunction of a particle that was initially in an $\hat{S}_z$ eigenstate will emerge from the final $\hat{S}_x$ measurement as a wavepacket that is split into two parts, corresponding to the two paths that would have been followed by particles initially in eigenstates of $\hat{S}_x$. It is only when we 'measure' the result of the experiment by allowing the particle to interact with a counter that the indeterminacy occurs. At this stage, only one of the two counters actually fires and the wavefunction becomes that corresponding to the appropriate eigenfunction of the operator corresponding to the measurement; this process is sometimes known as the 'reduction' or 'collapse' of the wavefunction. We are forced to

conclude that there are two different rules for calculating time dependence in quantum mechanics: the time-dependent Schrödinger equation which controls the deterministic evolution of the wavefunction, and the measurement postulate which produces the random indeterministic results of 'actual' measurements. Defining the boundary between these two processes constitutes the 'measurement problem' in quantum mechanics, to which we shall return in more detail later in this chapter.

Reality

Fundamentally, the aim of science is to describe nature as it really is. A technical word for such a description, which we shall use later on, is 'ontology'. For example, classical mechanics sets out the laws governing the relations between the positions and velocities of bodies in motion; these concepts are familiar to us and can be considered as 'real' attributes or properties of the bodies themselves. When the theory of electromagnetism was developed, the concept of reality was extended to cover the concept of fields. Although fields are less tangible than material objects, this idea was very successful in explaining a range of physical phenomena, including light and electromagnetic radiation, and now we have little trouble in thinking of classical fields as being part of reality.

Defining reality in the context of quantum mechanics is quite a different matter. We have emphasized throughout this book that the wavefunction is not to be considered as physical, but as a mathematical object from which the possible results of experiments and their relative probabilities can be deduced. By implication, the experimental results are themselves real; but experiments are performed on quantum objects, such as electrons, photons etc., and we appear to be eschewing the possibility of a realistic description of their properties. Moreover, as we shall see, even the reality of the experiments themselves can be difficult to maintain in a consistent way.

The first approach to the problem of quantum reality that we shall discuss consists of efforts made to go beyond the limitations of quantum mechanics and assign properties to quantum systems independently of their being observed. As we shall see, these theories give rise to conceptual difficulties of their own.

13.2 Hidden-variable theories

Many of the conceptual problems of quantum mechanics would be resolved if it could be shown that the predictions of the theory describe probabilities of actual occurrences, which may not themselves be directly observable. An analogy can be drawn with classical statistical mechanics where different possible events are ascribed particular probabilities, even though a detailed examination of the behaviour of the component atoms would show that these were following perfectly deterministic laws. If a similar deterministic substructure could be found to underlie quantum-mechanical indeterminism, this would have many

attractions. In the quantum case, however, we do not know what the substructure is or even if it exists at all. For this reason quantities postulated to underlie quantum behaviour, but which cannot be directly observed, are known as 'hidden variables' and theories based on such ideas are called 'hidden-variable theories'.

One necessary property of any hidden-variable theory is that it must reproduce the results of quantum mechanics in every case where these have been confirmed. Of course, if an as-yet-unperformed experiment could be devised in which the predictions of the new theory differed from those of quantum mechanics, then we should be in a position to make an experimental test to decide which of the two is correct. To date, all such tests have come down on the side of quantum mechanics, but if the opposite were ever found to be true, it would constitute one of the most important discoveries of modern physics.

de Broglie–Bohm theory

The earliest example of a hidden-variable theory was that proposed by de Broglie in 1927 as a direct consequence of his postulate of the existence of matter waves. This initial work has been developed and extended by a number of other workers since—notably David Bohm in the 1950s. For this reason it has become known as de Broglie–Bohm theory.

The basic idea behind de Broglie–Bohm theory is that atomic particles such as electrons always possess a real position and velocity. The wavefunction or 'matter wave' also exists and acts to guide the motion of the particles in such a way that their statistical properties are just those predicted by quantum mechanics. We can see how this works if we start from the Schrödinger equation governing the motion of a particle of mass m in one dimension under the influence of a potential $V(x)$:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V\psi \tag{13.1}$$

We define quantities R and S as real functions of x such that

$$\psi = R\exp(iS) \tag{13.2}$$

If we substitute from (13.2) into (13.1) we get

$$\begin{aligned} &i\hbar\frac{\partial R}{\partial t}\exp(iS) - \hbar R\frac{\partial S}{\partial t}\exp(iS) \\ &\quad = -\frac{\hbar^2}{2m}\left[\frac{\partial^2 R}{\partial x^2} + 2i\frac{\partial R}{\partial x}\frac{\partial S}{\partial x} + iR\frac{\partial^2 S}{\partial x^2} - R\left(\frac{\partial S}{\partial x}\right)^2\right]\exp(iS) + VR\exp(iS) \end{aligned} \tag{13.3}$$

We now cancel the common factor $\exp(iS)$ and separate the real and imaginary

parts to get two equations:

$$\left.\begin{aligned} \hbar\frac{\partial S}{\partial t} + \frac{\hbar^2}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V - \frac{\hbar^2}{2mR}\frac{\partial^2 R}{\partial x^2} = 0 \\ \hbar\frac{\partial R}{\partial t} + \frac{\hbar^2}{m}\frac{\partial R}{\partial x}\frac{\partial S}{\partial x} + \frac{\hbar^2}{2m}R\frac{\partial^2 S}{\partial x^2} = 0 \end{aligned}\right\} \tag{13.4}$$

If we multiply the second of (13.4) by R and rearrange it we get

$$\frac{\partial(R^2)}{\partial t} + \frac{\hbar}{m}\frac{\partial}{\partial x}\left(R^2\frac{\partial S}{\partial x}\right) = 0 \tag{13.5}$$

We can also rewrite the first of (13.4) as

$$\hbar\frac{\partial S}{\partial t} + \frac{\hbar^2}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V + Q = 0 \tag{13.6}$$

where Q is defined by

$$Q = -\frac{\hbar^2}{2mR}\frac{\partial^2 R}{\partial x^2} \tag{13.7}$$

Generalizing (13.5), (13.6) and (13.7) to three dimensions, we get

$$\frac{\partial(R^2)}{\partial t} + \frac{\hbar}{m}\boldsymbol{\nabla}\cdot(R^2\nabla S) = 0 \tag{13.8}$$

and

$$\hbar\frac{\partial S}{\partial t} + \frac{\hbar^2}{2m}(\nabla S)^2 + V + Q = 0 \tag{13.9}$$

where Q is now given by

$$Q = -\frac{\hbar^2}{2mR}\nabla^2 R \tag{13.10}$$

All we have done so far is to rewrite the Schrödinger equation in terms of the new functions R and S. However, we can now see how this leads to the physical ideas underlying de Broglie–Bohm theory. If, as the theory assumes, there are real particles that always have real positions and real velocities, the probability distributions for these positions should be subject to the classical continuity equation. In one dimension this means that the rate of change of the number of particles in a small region dx equals the net rate of flow of particles into the region. So, if $n(x,t)dx$ is the number in dx at time t,

$$\frac{\partial n}{\partial t} = (nv)_x - (nv)_{x+dx} = -\frac{\partial(nv)}{\partial x} \tag{13.11}$$

The three-dimensional equivalent of (13.11) is

$$\frac{\partial n}{\partial t} = -\boldsymbol{\nabla}\cdot(n\mathbf{v}) \tag{13.12}$$

If we now compare (13.11) and (13.12) with (13.25) and (13.8), we see that they will be equivalent provided R^2 is proportional to n and

$$\mathbf{v} = \hbar \nabla S/m \tag{13.13}$$

Because $R^2 = |\psi|^2$, the first condition is just the Born postulate relating the squared modulus of the wavefunction to the position probability density. It therefore follows that if we postulate that the particle velocity is given by the relation (13.13), the position probability distribution will evolve in time in exactly the way predicted by standard quantum mechanics.

To be consistent with quantum measurement theory (where, once a measurement has been made, the wavefunction corresponds to one of the eigenfunctions of the measurement operator) the statistical distribution must be transformed by a measurement into that corresponding with the new $|\psi|^2$. We note that this means that the eigenfunctions corresponding to the other eigenvalues will have no particles associated with them. In de Broglie–Bohm theory, such 'empty waves' exist, but play no part in determining the future behaviour of the particles.

Given this expression for the particle velocity, the second and third terms of (13.9) are now just the kinetic and potential energies of the particle. The fourth term, Q, however, has no classical analogue; according to de Broglie–Bohm theory, it is an additional potential known as the 'quantum potential'. It is a fundamental feature of the model that the difference between classical and quantum behaviour is due to this additional term. We can see how this comes about in the particular case of a conservative system where V is independent of time. It then follows from the separation of variables in the time-dependent Schrödinger equation that $\partial S/\partial t$ equals $-E/\hbar$, where E is the total energy of the system. Equation (13.9) is then just the classical conservation of energy equation, with the potential modified by the addition of the quantum potential. It can be shown that in the more general case where the potential, V, is time-dependent, (13.9) reduces to what is known as the Hamilton–Jacobi formulation of classical mechanics.

Although de Broglie–Bohm theory has traditionally been presented in terms of the quantum potential, more recent work in this area emphasizes the importance of (13.13) pointing out that all the results can be derived from this without invoking the idea of the quantum potential. After solving the Schrödinger equation to get ψ and hence S, we calculate the particle velocity using (13.13). Given this, we can straightforwardly compute the particle position as a function of time. Put this way, we can think of $\hbar \nabla S/m$ as analogous to a force that 'causes' a velocity $\mathbf{v}$. In this sense the ideas are somewhat reminiscent of the pre-Newtonian theories put forward by Aristotle.

Whichever way we interpret it in detail, de Broglie–Bohm theory seems to have succeeded. We have a realistic description of the properties of quantum objects, even when they are not being directly observed. If we include the quantum potential, we can simply extend the ontology that was developed for

classical physics into the quantum domain. We shall see later that things are not as simple as this, but we now proceed to illustrate the strengths of de Broglie–Bohm theory by considering its application to some examples.

Provided we can solve the Schrödinger equation, it is relatively easy to calculate the paths followed by the particles numerically. Given an initial position for the particle, we calculate its velocity using (13.13). This allows us to calculate the position after a short time dt; the new velocity is then obtained from the wavefunction at this position and time. This process can be continued until we have the full 'trajectory' followed by the particle, and can be repeated for different starting positions. We can often get a good idea of the expected behaviour without doing the full calculation and we are sometimes helped in this by a 'no-crossing rule'. This states that the de Broglie–Bohm trajectories can never cross each other. It follows directly from the fact that the wavefunction, including its phase, is a single valued, differentiable function of position. However, if the trajectories were to cross, there would have to be two possible values of $\mathbf{v}$ and hence ∇S at the crossing point.

We first consider the case where V is zero or uniform throughout space. We know that a solution of the Schrödinger equation in this case is a plane wave of constant amplitude (see chapters 2 and 3), so it follows from (13.9) that the quantum potential is constant everywhere. The particle therefore feels no force and moves with constant velocity, just as we would expect. This result can be generalized to the case of a 'wavepacket' where the wavefunction has the form

$$\psi(x) = f(x - pt/m)\exp[i(p_0x - E_0t/\hbar)] \tag{13.14}$$

where $f(x)$ is real and $E_0 = p_0^2/2m$. The function f is small when its argument is large, so that the wavepacket has the form of a pulse moving with velocity p_0/m. It follows directly from (13.13) and (13.14) that the particle moves with the same velocity as the wavepacket so that its position within it does not change.[3]

We now turn to the case where the particle passes through one or more slits in a screen. The wave is, of course, diffracted and the solutions to equations (13.8) and (13.9) have been solved numerically to produce trajectories for the particles in a number of such cases. An example of the trajectories associated with diffraction by a double slit is shown in figure 13.1, the starting positions being uniformly distributed across the slits. The statistical distribution of the particles arriving some distance from the slits shows the pattern of maxima and minima predicted by quantum mechanics and observed experimentally. However, the particles have followed clearly defined paths, just as they would in classical mechanics. We can therefore tell where a particle has been, though we cannot *predict* where it will go any more precisely than the Heisenberg uncertainty principle allows, because we cannot know *in advance* which part of which slit it started from.

[3] The exact solution of the Schrödinger equation in the case of a wavepacket is more complex as the various Fourier components evolve differently. However, (13.14) applies in the limit where the width of the wavepacket is large compared with the de Broglie wavelength, $\hbar/p_0$.

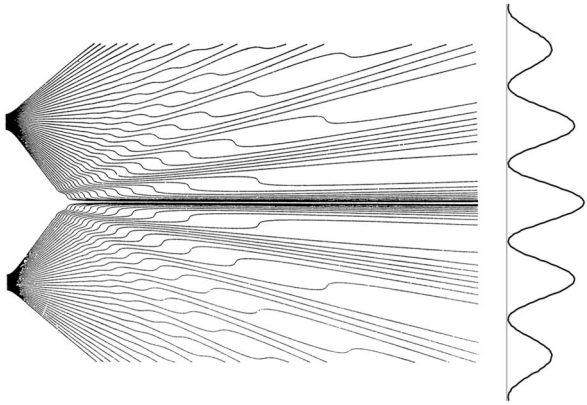


Figure 13.1. The particle trajectories calculated for double-slit diffraction using the de Broglie–Bohm hidden-variable theory. The trajectories cluster around the diffraction maxima in the expected way, as indicated by the Fraunhofer intensity pattern on the right. (Reproduced by permission from C. Philippides, C. Dewdney and B. J. Hiley, *Nuovo Cimento* B **52** 15–29 (1970). Copyright Società Italiana di Fisica.)

Figure 13.1 also illustrates the no-crossing rule mentioned earlier. As a result, all particles observed in the top half of the screen have passed through the upper slit and vice versa.

We turn now to the case of a particle confined to a potential well. This has been treated quantum mechanically in chapters 2 and 3, where we see that the ground-state wavefunction is usually real. It follows that S is zero everywhere in space so that $\nabla S = 0$ and the speed of the particle as described by de Broglie–Bohm theory must be zero. Thus in the ground state of a one-electron atom such as hydrogen the particle must be at rest at some unknown position. More complex behaviour emerges in states with non-zero orbital angular momentum, when the complex wavefunction is associated with the particle 'orbiting' the nucleus.

Overall then, de Broglie–Bohm theory appears to be so successful that the reader could be forgiven for wondering why it is not taken to be the orthodox interpretation of quantum mechanics and why it is not taught from the start, so avoiding many of the conceptual difficulties associated with the conventional approach.

Problems with de Broglie–Bohm theory

There are a number of reasons why de Broglie–Bohm theory has not gained universal acceptance.

First, there is the peculiar nature of the quantum potential itself. Potentials in physics, both classical and quantum, arise as a result of physical interactions between objects. For example, the potential experienced by an electron in a

hydrogen atom is due to the electromagnetic interaction between the negatively charged electron and the positive proton. But the quantum potential has no obvious physical basis: it arises purely out of the mathematics of the Schrödinger equation. Moreover, the forces associated with the theory do not obey Newton's third law relating to action and reaction: the wave influences the particle through the quantum potential, but the particle does not react on the wave.

Second, there are inconsistencies in the de Broglie–Bohm ontology. The theory began by considering the Schrödinger equation for a charged point particle, such as an electron. When this is in, say, the ground state of the hydrogen atom, quantum mechanics describes it as delocalized so that it interacts with another charged particle as if the charge were spread throughout the atom in proportion to $|\psi|^2$. The statistical results of de Broglie–Bohm theory averaged over many atoms are consistent with this but when applied to a single atom, the theory predicts that the particle is at rest at some unknown point. If the charge were localized at this point, the atom would possess an electric dipole moment, which could be measured. Experimentally, the dipole moment is zero, so the charge appears to be associated with the wavefunction rather than the particle, which contradicts our starting assumption. Similar arguments can be applied to the gravitational mass and many other physical properties. The 'particle' in de Broglie–Bohm theory seems to be a largely metaphysical object with no attributes apart from position!

Third, de Broglie–Bohm theory seems to make unnecessary postulates compared with conventional quantum mechanics. The particle trajectories can never be observed more accurately than the uncertainty principle allows, so the postulate that they exist can be seen as an unnecessary extravagance. The supporters of the theory, however, believe that this extravagance is preferable to the conventional approach which abjures any realistic description of the behaviour of quantum objects, and which has to make postulates such as wavefunction collapse that are unnecessary in de Broglie–Bohm theory.

The fourth and perhaps most important reason why de Broglie–Bohm theory is not generally accepted is that it is a 'non-local' theory. By this, we mean that a particle may be influenced not only by the potential at the point where the particle is, but also by its values at other points in space. This is not included in the formulation set out here, because we have restricted our consideration to the special case of a single particle moving in a potential $V(\mathbf{r})$. However, if we extend the argument to the case of two interacting particles, it follows from the discussion in chapter 10 that the wavefunction, and hence the quantum potential, are functions of the coordinates of both particles. They therefore do not exist in real space but in an abstract six-dimensional 'configuration space' spanned by the six coordinates. An appropriate change in the 'real' potential $V(\mathbf{r}_1, \mathbf{r}_2)$ in the vicinity of one particle produces an immediate change in the joint wavefunction and hence the quantum potential affecting the other particle—even though the latter may be a large distance from the former.

Scientists have always been very reluctant to accept non-locality as part of a physical theory. The idea of being able to change the behaviour of a system

some distance away without communicating with it directly and without using some kind of field to transmit the influence seems more like magic than physics. Indeed, when Newton first proposed his law of gravitation, he suggested that one body was able to exert an 'action at a distance' on another. This considerably delayed the general acceptance of Newton's ideas, and was finally overcome only when the idea of a gravitational field was proposed which transmitted the force between the bodies and then acted locally. More recently, the theory of relativity has shown that all fields capable of transmitting information must do so no faster than the speed of light, which strengthens our belief in the reality of the classical field.

If, despite all this, we accept de Broglie–Bohm theory despite its non-locality, all the results of quantum theory can indeed be reproduced. However, as we shall see later, serious problems arise when we try to reconcile this non-local theory with the principles of relativity.

If de Broglie–Bohm theory is not acceptable, might there be another form of hidden-variable theory that could 'explain' quantum mechanics without having this disadvantage? In the 1950s this question very much interested the quantum physicist John Bell, who believed that quantum mechanics should have a realistic substructure and that de Broglie–Bohm theory had been 'scandalously neglected'. Ironically, he found that it was impossible for any local hidden-variable theory ever to reproduce all the results of quantum mechanics. Because this result is so important and has been so influential over the development of thought in this area, we devote the next section to a discussion of it.

13.3 Non-locality

The non-local implications of quantum mechanics were first discussed by A. Einstein, B. Podolski and N. Rosen in a paper published in 1935, and work in this area is often referred to as the 'EPR problem' after the initials of their surnames. The problem, though not its solution, was greatly clarified by an illustrative example proposed by David Bohm in the 1960s. This consists of a pair of spin-half particles, each of which is known to have zero orbital angular momentum and whose total spin is also known to be zero. Such a system can be created if a beam of low-energy protons undergoes s-wave scattering from hydrogen gas; as we saw at the end of chapter 10, the incident and target protons move apart in a state where the total orbital angular momentum and the total spin of the pair are both zero. However, the details of how this might be done are not important, and a system with the same properties that is easier to prepare consists of two photons in a state where their plane polarizations are mutually perpendicular—see later. Our discussion will concentrate on the spin case, although we shall discuss experiments based on the observation of polarization. However they are created, the particles are allowed to move apart until they are widely separated. We then measure (say) the z component of spin of one particle

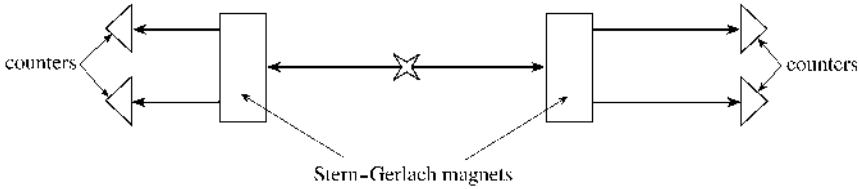


Figure 13.2. An example of non-locality. A pair of spin-half particles is created in a state with zero total spin and the spin of each is analysed by a Stern–Gerlach magnet. Quantum mechanics says that a measurement on one of the particles collapses the whole wavefunction.

($\hat{S}_{z1}$) and follow this by measuring the same component of the other ($\hat{S}_{z2}$) (cf. figure 13.2). Because the total spin is zero, we might expect equal and opposite results ($\pm\frac{1}{2}\hbar$), and this is indeed what is observed.

This result is not as obvious as it may appear at first sight. Remember that when we make a quantum measurement, we cause the system to collapse randomly into a state corresponding to one of the measurement outcomes. Thus, measuring the spin of one particle will produce, say, a positive result and a collapse of its wavefunction. However, if the spin of the other is certain to be negative, the whole wavefunction must have collapsed, not just the part associated with the first particle measured. This is indeed what quantum mechanics predicts: the wavefunction of two particles with opposite spin has the form

$$\psi(1,2) = 2^{-1/2}[\alpha(1)\beta(2) - \beta(1)\alpha(2)] \tag{13.15}$$

This expression is independent of the absolute direction represented by α and β (see the discussion of entanglement in chapter 12). Following standard quantum measurement theory, if a measurement of the spin of particle 1 yields (say) a positive result, this will lead to a collapse of the whole wavefunction into the state $\alpha(1)\beta(2)$. A subsequent measurement of the spin of particle 2 *must* then produce a negative result.

These predictions of quantum mechanics are confirmed experimentally, as we shall discuss in more detail later, but the non-local implications should be clear. The two particles may be well separated (by many metres in some experiments), but a measurement on one of them changes the wavefunction associated with the other. In fact the second measurement is unnecessary: we know exactly what its consequences will be as soon as the first measurement has been made, *even though the two particles are widely separated and there is no known interaction between them.*

Consider a classical analogy. If an object with zero angular momentum splits into two parts and separates, measurement of the angular momentum of one part tells us that of the other will be equal and opposite. However, in the

classical case the act of measurement has no effect on the system, while the opposite is true in the quantum case. It is tempting then to believe that a system like a spin-half atom is more classical than we had thought and to look for an explanation similar to the classical description. Such a model would, in fact, be a local hidden-variable theory (LHVT) because the result of a classical angular momentum measurement would be determined by the actual angular momentum carried away by each fragment separately. Can we devise a theory in which the results of spin measurements would be similarly determined by hidden variables localized in each spin separately?

Any such LHVT would have to successfully predict the results of any feasible measurement that could be made on the two-particle system, and not just the case where the measurement direction is the same on each side. It turns out that this is not possible, as was proved by John Bell in 1964, and we shall set out a proof of Bell's theorem shortly. However, before we can proceed to this, we will need the quantitative predictions of quantum mechanics for a measurement of the component of the second spin in a direction at an angle θ to the z axis ($S_{\theta 2}$) following a measurement of the z component of spin of the first particle (S_{z1}). We shall assume that all angular momenta are expressed in units of $\frac{1}{2}\hbar$ so that the result of any such measurement is either $+1$ or -1. In the case where the result of the first measurement is positive, we conclude that S_{z2} is consequently negative and that the spin part of the wavefunction of this particle can be represented by the column vector $\beta_z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, as in table 6.1. Taking the plane containing the two spin directions as the xz plane, the eigenvectors of $S_{\theta 2}$ are given in chapter 6 as

$$\alpha_\theta = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} \qquad \text{and} \qquad \beta_\theta = \begin{bmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix}$$

when the eigenvalues are positive and negative respectively. We can now use the measurement postulate to expand the initial wavevector as a linear combination of the eigenvectors:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \sin(\theta/2) \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} + \cos(\theta/2) \begin{bmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix} \tag{13.16}$$

The probability of the result of the second measurement also being positive is therefore $P_{++}(\theta)$ where $P_{++}(\theta) = \sin^2(\theta/2)$ and that of it being negative is $P_{+-}(\theta) = \cos^2(\theta/2)$. Similar arguments lead to expressions for the similarly defined probabilities $P_{-+}(\theta)$ and $P_{--}(\theta)$, and we have

$$\left.\begin{aligned} P_{++}(\theta) &= \sin^2(\theta/2) \\ P_{+-}(\theta) &= \cos^2(\theta/2) \\ P_{-+}(\theta) &= \cos^2(\theta/2) \\ P_{--}(\theta) &= \sin^2(\theta/2) \end{aligned}\right\} \tag{13.17}$$

Bell's theorem

Bell's theorem states that it is impossible for any local hidden-variable theory (LHVT) to reproduce all the predictions of quantum mechanics.

To prove Bell's theorem we consider the constraints that any LHVT must impose on the results of measurements on the spin components of a spin-half atom, and which can be tested by measurements on entangled pairs of particles, such as those discussed earlier. According to our LHVT, the result of any spin measurement is determined in advance by some hidden variable contained in the particle. We consider the components of spin in three directions (1,2,3) (which need not be mutually perpendicular). A set of N such atoms will contain a subset of $n(1_+, 2_+, 3_+)$ particles, each of which would yield a positive result for a measurement of spin in any of the three directions; $n(1_+, 2_+, 3_-)$ particles where a measurement in direction 1 or 2 will yield a positive result, with a negative result if the measurement is made in direction 3 etc. The set of N particles is therefore composed of eight non-overlapping subsets defined by the signs of the three spin components. We cannot identify which atom is in which subset, because we can measure only one spin component on each atom without affecting the others, but if the LHVT is true all the particles must belong to one or other of the eight subsets.

Suppose now each atom is one of an entangled pair and we make a measurement of, say, spin component 1 on one member of each pair and a different component (say 2) on the other. *If the system is subject to a LHVT, these measurements cannot affect each other.* Hence, because the total spin is zero, a particular result on particle 2 means that we would have obtained the opposite result on particle 1 for that component. We have therefore found a way of measuring two spin components on particle 1 while only disturbing it once. That is, we can measure the number of particles with positive spin in directions 1 and 2; we call this $n(1_+, 2_+)$, with corresponding quantities for the other possible signs.

Given this it follows that

$$\begin{aligned} n(1_+, 2_+) &= n(1_+, 2_+, 3_+) + n(1_+, 2_+, 3_-) \\ n(1_+, 3_+) &= n(1_+, 2_+, 3_+) + n(1_+, 2_-, 3_+) \\ n(2_-, 3_+) &= n(1_+, 2_-, 3_-) + n(1_-, 2_-, 3_+) \end{aligned} \tag{13.18}$$

with similar results for the other five possible measurement outcomes. We note that, even given a LHVT, it is not possible to measure more than one of the quantities on the left-hand side of (13.18) on a particular set of N particles. However, if we use three similarly prepared sets and if N is large, the results should be the same as we would have obtained if we had made the measurement on any one of the three sets.

It follows directly from (13.18) that

$$n(1_+, 2_+) - n(1_+, 3_+) + n(2_+, 3_-) = n(1_+, 2_+, 3_-) + n(1_-, 2_-, 3_+)$$

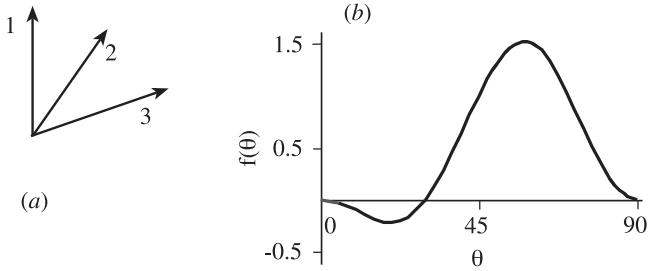


Figure 13.3. (*a*) The measurement directions and the angles between them. (*b*) The left-hand side of (13.22) as a function of θ. The fact that it is negative for $\theta > 60°$ proves Bell's theorem.

which means that

$$n(1_+, 2_+) - n(1_+, 3_+) + n(2_+, 3_-) \geqslant 0 \tag{13.19}$$

This is a version of what is known as *Bell's inequality*.

We now compare this result with the predictions of quantum mechanics. Referring to (13.17), we have $n(1_+, 2_+) = NP_{+-}(\theta_{12})$ where θ_{12} is the angle between directions 1 and 2. Thus, (13.19) becomes

$$P_{+-}(\theta_{12}) - P_{+-}(\theta_{13}) + P_{++}(\theta_{23}) \geqslant 0 \tag{13.20}$$

or

$$\cos^2 \frac{\theta_{12}}{2} - \cos^2 \frac{\theta_{13}}{2} + \sin^2 \frac{\theta_{23}}{2} \geqslant 0 \tag{13.21}$$

for all possible values of the three angles.

To prove Bell's theorem we only have to show that (13.21) is false for some particular values of the angles. We consider the case where all three measurement directions are in the same plane and in the configuration shown in figure 13.3(*a*). In this case, $\theta_{12} + \theta_{23} = \theta_{13}$ and we specialize further by putting $\theta_{13} = 3\theta_{12}$ while writing $\theta_{12}/2$ as θ. This leads directly to

$$\cos^2 \theta + \sin^2 2\theta - \cos^2 3\theta \geqslant 0 \tag{13.22}$$

This function is plotted in figure 13.3(*b*) where we see that, although it is positive for θ greater than about 30°, it is negative for larger values of the angle. In particular, for $\theta = 20°$ the expression has the value -0.22.

We have therefore shown that there is a clear contradiction between the predictions of any LHVT and those of quantum mechanics, and we have proved Bell's theorem.

Example 13.1 A 'real-spin' hidden-variable theory As an example of the application of Bell's theorem, we consider a particular theory in which we try

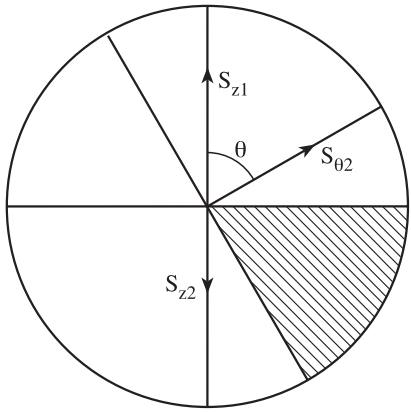


Figure 13.4. According to the 'real-spin' hidden-variable theory, the number of particle pairs where the z component of the spin of the first particle and that of the second particle in the direction defined by θ are both positive is proportional to the shaded area.

to adapt the classical theory of angular momentum to the quantum situation. Because this is to be a local theory it is necessarily different from the de Broglie–Bohm theory discussed earlier. We first postulate that all the components of angular momentum of a particle always have definite, though unknown, values. It follows that, as all the components of the total combined spin of the two particles are zero, the 'real-spin' vectors associated with the two individual particles are equal and opposite and remain so as the particles separate. In quantum mechanics only one component of spin can be measured without affecting the other two and we shall assume that this property holds in our model. A further assumption has to be made concerning the interaction between the 'real spin' and the measuring apparatus if the only possible results of the measurements are to be plus or minus one (in units of $\frac{1}{2}\hbar$): we postulate that if any spin component is 'really' positive (or negative) then a measurement will always yield the result $+1$ (or -1) whatever the actual magnitude of the component.

It follows directly from this that, if S_{z1} is measured as $+1$, then the z component of the second spin must be negative and the 'real' spin vector of the latter must lie somewhere on a hemisphere whose symmetry axis is the negative z axis (see figure 13.4). Moreover if the component $S_{\theta 2}$ is also positive, the corresponding true spin must lie somewhere in a hemisphere whose symmetry axis makes an angle θ with z. Given that the absolute orientation in space of the two spin vectors is random—provided of course that they are always equal and opposite to each other—the probability, $P_{++}(\theta)$, that both components are positive is proportional to the volume of overlap of the two hemispheres. We see from figure 13.4 that this is, in turn, proportional to θ. As we know that when $\theta = \pi$ the result must be a certainty, that is $P_{++}(\pi) = 1$, the constant of

proportionality must be equal to $1/\pi$. The other probabilities can be estimated in a similar manner so we have

$$\left.\begin{aligned} P_{++}(\theta) &= \frac{\theta}{\pi} \\ P_{+-}(\theta) &= 1 - \frac{\theta}{\pi} \\ P_{-+}(\theta) &= 1 - \frac{\theta}{\pi} \\ P_{--}(\theta) &= \frac{\theta}{\pi} \end{aligned}\right\} \tag{13.23}$$

Substituting these expressions into (13.20) we get

$$\theta_{13} + \theta_{23} - \theta_{12} \geqslant 0 \tag{13.24}$$

This is true for all possible values of the angles: the left-hand side of (13.24) is zero when all directions are in the same plane and direction 3 lies between directions 1 and 2. Thus Bell's inequality is satisfied. However, the probabilities clearly differ from the expressions given in (13.17), and our theory does not agree with quantum mechanics.

Experiments

As quantum mechanics has been so well verified over a wide range of phenomena, it might be thought that any further experimental test of the discrepancy between its predictions and those of hidden-variable theories would be unnecessary. However, it turns out that, because of the particular nature of the correlation discussed earlier, none of the experiments that had been performed before the formulation of Bell's theorem in 1969 provided a direct test of this point. In fact the form of Bell's inequality set out here cannot be tested directly as it assumes that all particle pairs have been detected; while, in any experiment conducted with real detectors, some particles are always missed. However, other versions of Bell's theorem have been derived that are not subject to this criticism. Moreover, the proofs have been extended to include hidden-variable theories that are not perfectly deterministic, but include an element of randomness in determining their predictions.

As previously mentioned, the polarization of a photon follows quantum rules that are essentially the same as those governing the spin of a spin-half particle: photon polarization can be measured as being directed parallel or perpendicular to some direction in space and such measurements follow rules that are identical to those applying to the measurement of spin as 'up' or 'down'. Some of the earliest definitive experiments were performed by Alain Aspect in France in the early 1980s and figure 13.5 shows the essentials of his apparatus. Pairs of photons are emitted from the centre of the apparatus in a quantum state such that the

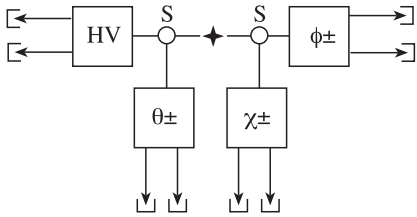


Figure 13.5. The Aspect experiment to compare the predictions of quantum mechanics with those of Bell's theorem for pairs of polarized photons. The photons can be switched into either channel on each side, so that measurements can be made with four different pairs of polarizations. The switches marked S are changed at a rate faster than the time required to send a signal from one side of the apparatus to the other at the speed of light.

polarizations of the members of each pair are always perpendicular (just as the spins of the atom pairs discussed earlier are opposite). Each photon is then directed into one of two polarizers that determines whether its spin is parallel or perpendicular to some direction. Which polarizer a particular photon enters is determined by a switch that operates at a frequency of around 10^8 Hz. The polarizers are oriented so that one of the left-hand polarizers is set to measure polarization as either horizontal or vertical (H/V) and the other as either parallel or perpendicular to a direction at an angle θ to the horizontal, while those on the right-hand side are set at ϕ and χ to the horizontal respectively. After a number of photon pairs have passed through the apparatus, we can determine the number which have been measured as, say, vertical on the left-hand side and parallel to ϕ on the right, which we call $n(V, \phi_+)$. Other pairs are similarly denoted. Separate runs are performed in which the left-hand photon is detected with the polarizers removed on that side and the right-hand photon is detected with its polarization parallel to θ: the number detected in this way is denoted as $n(\theta_+)$. The extended version of Bell's theorem applicable to this situation is

$$n(V, \phi_+) - n(V, \chi_+) + n(\theta_+, \phi_+) + n(\theta_+, \chi_+) \leqslant n(\theta_+) + n(\phi_+) \qquad (13.25)$$

It can be shown that the maximum disagreement with the predictions of quantum mechanics for this set-up occurs when the angles are $\theta = 45°$, $\phi = 67.5°$ and $\chi = 22.5°$; the left-hand side should be larger than the right by $0.112N$, where N is the total number of photon pairs recorded in each run. The experimental result in this configuration was $0.101N$, which agrees with quantum prediction within experimental error, but is completely inconsistent with Bell's theorem.

The Aspect experiment has another property that is of considerable significance. This is that, because the switches operate at such a high rate, the time between switchings is considerably shorter than the time needed for light to travel from one side of the apparatus to the other. Thus, if we were to imagine that the correlations were established as a result of some unknown physical interaction

between one side of the apparatus and the other, then this would have to be propagated faster than light. Overall then, the Aspect experiment seems to provide conclusive proof that a correct hidden-variable theory must be non-local in nature and any non-local influence must be propagated through space instantaneously, or at least at superluminal speed. This is certainly the conclusion drawn by nearly all workers in the field, but it should be noted that the argument is not quite rigorous. This is because the proof of the extended Bell's theorem relies on several additional assumptions, among which is the apparently reasonable one that polarizers may attenuate, but can never amplify, light. It has been pointed out that, because the efficiency of photodetectors is quite low, it is possible to postulate a breach of this 'no enhancement' postulate without implying any overall increase in the energy of the system, and so to produce a LHVT that is compatible with the experimental results obtained so far. However, the majority view is that, although this loophole does exist, it is very small and will eventually be closed when high-efficiency photo-detectors are developed.

In the twenty or so years since the Aspect experiment, Bell's theorem (with the additional assumptions mentioned above) has been re-tested in a number of experimental situations. By the end of the twentieth century, EPR-type correlations had been observed between pairs of photons separated by several kilometres, experiments had been performed with random switching of the outputs (in contrast to the periodic switching used in the Aspect experiment) as well as on photon triplets. In every case, the quantum predictions were confirmed within experimental error and some form of Bell's inequality was clearly breached.

An important feature of all correlations between separated photons is that they cannot be used by themselves to transmit information—i.e. signalling is forbidden. To see this, imagine observing the results of measurements made on one side of an apparatus such as that shown in figure 13.2. The spin would be found to be positive or negative at random, whatever the setting of the apparatus on the other side. Thus there is no way for an experimenter on one side to signal to the other merely by changing the orientation of her apparatus. It is only when the records of the two sets of observations are brought together and compared that the correlations are manifested. Thus, EPR non-locality does not breach the fundamental principles of relativity.

Accepting that Bell's theorem and the Aspect experiment rule out LHVTs, we now discuss how non-local de Broglie–Bohm theory models the case of entangled spin-half particles. When the two particles emerge from the source, they are represented by wavepackets, associated with each of which is a Bohm particle. The rule requiring the statistical distribution of Bohm particles to be $|\psi|^2$ ensures that the positions of the particles within the wavepackets are not correlated with the spin directions. What happens in a spin measurement is that, as a wavepacket passes through a Stern–Gerlach (SG) apparatus, it splits into two components with opposite spin. Only one of these contains the Bohm particle, which is subsequently detected.

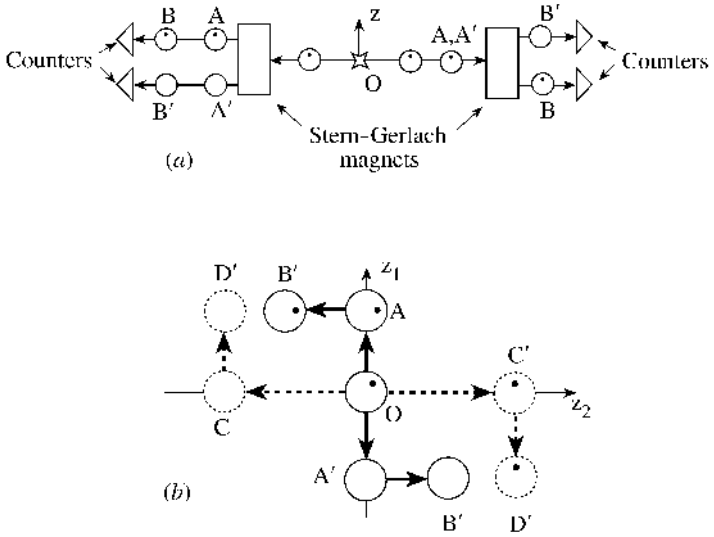


Figure 13.6. (*a*) A pair of wavepackets (open circles) containing Bohm particles (filled circles) in their upper halves are split by Stern–Gerlach magnets oriented to measure the z component of spin. The figure shows the positions of the packets and particles at several stages of evolution. (*b*) The unbroken lines illustrate the same process in the z_1/z_2 plane of the two-particle configuration space, while the broken lines show what would happen if the right-hand magnet were the closest to the source.

Returning to the two-particle case, we will assume for simplicity that the same (z) component of spin is measured in each half of the apparatus (see figure 13.6(*a*)). To agree with quantum mechanics and experiment, the left- and right-hand Bohm particles must end up in opposite component packets. It is helpful to represent the motion of the two particles by that of a single point in the six-dimensional 'configuration space' defined by the coordinates of the two particles. As the wavefunction is a function of all these coordinates (cf. chapter 8) it must be a differentiable single-valued function of all of them, and the de Broglie–Bohm no-crossing rule holds in configuration space. We are primarily interested in motion in the z direction as defined in figure 13.6(*a*), and the projection of configuration space onto the two-dimensional plane defined by z_1 and z_2 is shown in figure 13.6(*b*). Remembering that the particles are randomly distributed with a probability distribution $|\psi|^2$, around 25% of the pairs will have both Bohm particles in the top (i.e. $z > 0$) halves of their wavepackets. This corresponds to the representative points in configuration space being in the top right-hand quadrant of the wavepacket—represented by the circle near the origin of figure 13.6(*b*). We first consider the case (represented by continuous lines in figure 13.6(*b*)) where the left-hand SG apparatus is nearer the source of

the particles than the right-hand one. As the wavepacket interacts with the left-hand apparatus, it splits into two and the Bohm particle joins the upward moving packet. When the right-hand SG is reached (points A and A′ in configuration space) the correlation between the spin directions ensures that the wavepackets move to the points B and B′. We note that the Bohm particle can follow the wavepacket from A to B, despite being on the 'wrong' side of the wavepacket, without violating the no-crossing rule in configuration space. When the Bohm particles are subsequently detected, we conclude that the left- and right-hand spins are positive and negative respectively.

Suppose, now, that we had made the measurement on the right before that on the left. The configuration-space point would have followed the route OC′ D′ and we would have concluded that the left- and right-hand spins were negative and positive respectively—directly opposite to the previous result. In both cases, provided we start with the particles randomly oriented in the wavepackets, the statistical outcome would be the same and would agree with quantum-mechanical predictions.

The purpose of de Broglie–Bohm theory is not only to predict results that we can just as well calculate by quantum mechanics, but also to provide a realistic ontology. The Bohm particles are supposed to be 'really there'! Now remember that the Aspect experiment has shown that spin correlations persist even when the time between the measurements is less than the time for light to travel across the apparatus. However, in such a case the theory of relativity allows us to reverse the time ordering of events simply by observing them from an inertial frame that is moving relative to the laboratory. It follows that the destinations of the Bohm particles seem to depend on which frame of reference we make our observations in!

However, there can only be one measurement outcome and this must be the same for both observers. The only things that might differ are the conclusions the observers reach about the properties of the Bohm particle. The observer who sees, say, the positive result on particle 1 first, concludes that this particle must be somewhere in the upper half of its wavepacket. In contrast, the observer who sees the negative result on particle 2 first, concludes that this particle is in the lower half of its wavepacket. Nevertheless, we have shown that the theory requires there to be some pairs in which both particles are in the upper halves. We can resolve all this if we are prepared to accept that a description of reality, as reflected in the particle positions, is relative to the observer, but this greatly weakens the objectivity that the theory was supposed to provide. Alternatively, the theory of relativity might not apply to the motion of the Bohm particle, which occurs in some privileged frame of reference such that time ordering is the same for all observers. A third possibility is that counters do not detect Bohm particles, but are activated by the wavefunction, even if it is 'empty', but then the main gain of Bohm theory, which was to provide a deterministic ontological substructure for quantum measurements, would be lost.

Combined with the problems of non-locality and the criticisms set out earlier,

these arguments must cast considerable doubt on the correctness and ontological usefulness of the Bohm model.

If we reject non-local theories such as that of de Broglie–Bohm, what is the alternative? One possibility would be to look for another hidden-variable theory that was not subject to the criticisms set out above. Alternatively we could return to quantum mechanics and look again at the standard interpretation. An important feature of this is that properties which, in principle, cannot be measured in a particular experimental situation do not have any real existence. Thus, if we pass a spin-half particle through a Stern–Gerlach apparatus oriented to measure $\hat{S}_z$, then the other spin components are not just unmeasurable but have no reality. The traditional, 'Copenhagen', interpretation of quantum mechanics, propounded by Niels Bohr in particular, assigns physical properties not to individual quantum systems on their own, but to quantum systems in conjunction with measuring apparatus; the only physical properties that are real in this situation are those that can be measured by the apparatus. The proof of Bell's theorem then fails because the unmeasured spin components do not have values that are functions of the hidden variables: they just do not exist. Clearly, however, such an interpretation of quantum mechanics implies that we know what a measurement is, and this brings us back to the measurement problem.

13.4 The quantum-mechanical measurement problem

Probably the most difficult and controversial conceptual problem in quantum mechanics concerns the nature and meaning of the quantum theory of measurement described in chapter 4. The relevant theory is contained in postulates 4.2, 4.3 and 4.4 which state that the measurement of a physical quantity always produces a result equal to one of the eigenvalues (q_n) of the operator ($\hat{Q}$) representing that quantity; that the wavefunction immediately after the measurement is the same as the corresponding eigenfunction (ϕ_n); and that, if the wavefunction is ψ before the measurement, the probability of obtaining the result q_n is equal to $|c_n|^2$ where $\psi = \Sigma_n c_n \phi_n$. The effect of the measurement is therefore to cause the wavefunction to be changed from ψ to ϕ_n. We can represent this process by

$$\psi \underset{\substack{\text{measurement}\\ \text{giving}\\ \text{result } q_n}}{\longrightarrow} \phi_n \tag{13.26}$$

This process is known as the 'reduction' or 'collapse' of the wavefunction. It is fundamentally a 'stochastic' process which means that the actual outcome is unpredictable, although the statistical properties of a large number of similar experiments can be calculated.

In contrast, postulate 5 states that the evolution of the wavefunction *between measurements* is governed by the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi \tag{13.27}$$

This is a linear equation, which means that the wavefunction at time t is entirely determined by the wavefunction at $t = 0$, along with the Hamiltonian $\hat{H}$, and there is no randomness. Thus, there are two different time dependencies in quantum mechanics: 'unitary evolution' under the action of the time-dependent Schrödinger equation and 'collapse' associated with a measurement. Although it is always easy in practice to decide which process occurs in any particular case, defining objective criteria for this is much more difficult.

To look more closely at what constitutes a measurement we consider once again the example of the measurement of a component of spin of a spin-half particle using an SG apparatus. If the component to be measured is S_z, whose eigenvalues are $\pm\frac{1}{2}\hbar$ and whose eigenvectors are represented by α_z and β_z, and if the initial spin state of the particle is α_x, (13.26) becomes

$$\left.\begin{aligned} \alpha_x = \frac{1}{\sqrt{2}}(\alpha_z + \beta_z) &\underset{\substack{\text{yielding}\\ S_z=\frac{1}{2}\hbar}}{\longrightarrow} \alpha_z \\ \alpha_x = \frac{1}{\sqrt{2}}(\alpha_z + \beta_z) &\underset{\substack{\text{yielding}\\ S_z=-\frac{1}{2}\hbar}}{\longrightarrow} \beta_z \end{aligned}\right\} \qquad (13.28)$$

It is important to note that this measurement, and the subsequent collapse of the wavefunction, are not achieved simply by passing the particles through an appropriately oriented SG magnet. This follows from a consideration of the arrangement illustrated in figure 13.6 where particles with a known S_x are passed through such a magnet oriented to 'measure' S_z and then directed back into a common path. Provided their path lengths are carefully controlled, the waves associated with the two paths interfere to re-construct the initial wavefunction. This is possible because no information has been obtained about the value of S_z. Thus, the wavefunction has not collapsed and its spin part is still α_x after the process. It follows that, in order to make successful measurements of S_z, some detecting device or counter must be introduced to record through which channel the particles passed. It is the presence of such recording apparatus which apparently causes the reduction of the wavefunction described by (13.26) and (13.28).

The description of the quantum theory of measurement in the previous paragraph would be perfectly correct and sufficient, provided we could treat the recording equipment as a separate piece of apparatus obeying the laws of classical physics. However, quantum mechanics is believed to be a universal theory, capable of describing macroscopic as well as microscopic objects, so why should we not be able to treat the whole set up, including the detector, as a quantum system subject to the time-dependent Schrödinger equation? Let us see what happens if we do. We again consider the measurement of $\hat{S}_z$ described earlier, but now assume that the passage of the particle through the positive channel of the SG magnet is recorded by some detector, and that this detector can itself be described

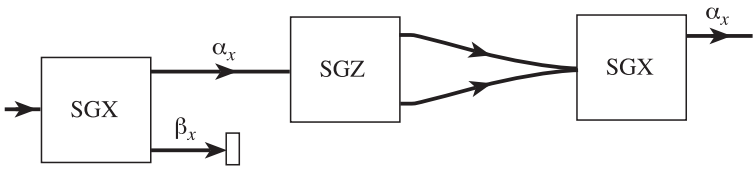


Figure 13.7. Spin-half particles with positive x component of spin pass through an SGZ apparatus and are then directed along a common path into an SGX without any record of their z component having been made. A further measurement of their x component invariably produces a positive result.

by a wavefunction which can take one of two forms: χ_0 before the particle is detected and χ_+ if $S_z = \frac{1}{2}\hbar$.

We consider first the case where the initial spin state of the particle is represented by α_z; the total wavefunction of the whole system consisting of particle and measuring apparatus is then

$$\psi_0 = \alpha_z \chi_0 \tag{13.29}$$

After the particle has passed, χ_0 will be changed to χ_+ and the total wavefunction becomes

$$\psi_+ = \alpha_z \chi_+ \tag{13.30}$$

However, if the initial state of the particle is β_z nothing is detected and the total wavefunction is still

$$\psi_0 = \beta_z \chi_0 \tag{13.31}$$

What we really want to know is what happens when the particle is not in an eigenstate of $\hat{S}_z$ before the measurement, but has a wavefunction whose spin part is, say, α_x. The initial state of the whole system, particle-plus-detector, is then

$$\psi_0 = \alpha_x \chi_0 = \frac{1}{\sqrt{2}}(\alpha_z + \beta_z)\chi_0 \tag{13.32}$$

Now, if the time evolution of the wavefunction is governed by the time-dependent Schrödinger equation, each term on the right-hand side of (13.32) evolves in exactly the same way as it does when the initial wavefunction is the appropriate eigenfunction of $\hat{S}_z$. Hence, using (13.30), (13.31) and (13.32), we find that the total wavefunction after the measurement is

$$\psi = \frac{1}{\sqrt{2}}(\alpha_z \chi_+ + \beta_z \chi_0) \tag{13.33}$$

States like the right-hand side of (13.33) are called 'entangled' states because the spin and counter states are entangled together in a linear combination. As we might have expected, (13.33) contradicts (13.28) which implies that the

wavefunction should have collapsed into *either* $\alpha_z\chi_+$ *or* $\beta_z\chi_0$ as a result of the measurement. Thus, if the detector were to contain a pointer that could occupy one of two positions depending on which of these states it was in, then (13.33) would imply that the pointer was somehow delocalized between two positions. Although predicted by quantum mechanics, it would seem quite contrary to our everyday experience if classical macroscopic objects such as detectors were to be in such a state.

We might try to resolve the problem by postulating that no collapse occurs until a measurement is made on the whole quantum system, particle-plus-detector. We could then use further apparatus to measure its state, when its wavefunction might collapse to $\alpha_z\chi_+$ or $\beta_z\chi_0$, depending on the result of this measurement. But this new measuring device could of course be considered as part of an even larger system which would then be isolated and whose wavefunction would again have a form similar to that of (13.33). There is therefore a potentially infinite chain of measurements and there would seem to be no point at which we can unambiguously state that the wavefunction has collapsed.

Schrödinger's cat

The apparently paradoxical results discussed earlier are vividly illustrated by an example first discussed by Schrödinger in the early days of the subject. In this thought experiment, Schrödinger imagined the result of a quantum-mechanical measurement (such as obtaining the value $\frac{1}{2}\hbar$ from a measurement of S_z on a particle whose initial state was α_x) being used to fire a gun (or trigger some other lethal device) in the direction of an unfortunate cat which is consequently killed. If, however, the result is $-\frac{1}{2}\hbar$, the gun is not fired and the cat remains alive. The whole apparatus is enclosed in a box, which is opened at a later time when the state of the cat (live or dead) is examined, from which observation the value of S_z can be deduced. By analogy with the earlier discussion, the time-dependent Schrödinger equation predicts that the wavefunction of the box and all its contents will be given by (13.33) where χ_+ and χ_0 now describe the whole apparatus (apart from the particle, but including the cat) when the cat is alive and dead respectively. But this would imply that before the box is opened the state of the cat is neither alive nor dead! If we reject this apparently absurd conclusion we must ask at what point the wavefunction collapses: is it when the particle enters the magnet (clearly not for reasons given here), when the gun fires, when the cat dies or when?

Decoherence

There is one important difference between a state described by a linear combination of spins as in (13.28) and a similar linear combination involving macroscopic measurement apparatus such as (13.33). Essentially this lies in the complexity of the many-body wavefunction representing a macroscopic object such as a measuring apparatus or a cat. This which ensures that it

is, in practice, completely impossible to perform an interference experiment to reconstruct the state $\alpha_x \chi_0$ from the linear combination $2^{-1/2}(\alpha_z \chi_+ + \beta_z \chi_0)$. If such a reconstruction is impossible, the results of all subsequent measurements will be the same as they would be if the system had 'really' collapsed into an eigenstate after the particle has been detected. The process whereby the macroscopic systems evolve into a state where interference is for all practical purposes forbidden is known as *decoherence*.

To see in more detail how decoherence works, we use a slightly generalized form of (13.33)

$$\psi = A\alpha_z \chi_+ + B\beta_z \chi_0 \tag{13.34}$$

We assume the χs are normalized and that $|A|^2 + |B|^2 = 1$. Consider the statistical results of a large number N of identical SGZ measurements, including appropriate recording apparatus, whose wavefunctions have the form (13.34). Such a collection of identical experiments is known as an *ensemble* and the Schrödinger equation predicts that, after the measurement, the ensemble will be in a so-called 'pure state'. In contrast, consider the case of N particles where $N|A|^2$ are known from the start to be in the state α while $N|B|^2$ are in the state β. Such an ensemble is known as a *mixture* and, in this case, measurements of $\hat{S}_z$ simply reveal properties the particles already possess. In the context of ensembles, therefore, collapse corresponds to a transition from a pure state to a mixture.

How could we find whether a given ensemble is in a pure or a mixed state? Consider an operator $\hat{Q}$ which represents some physical operation on the whole system of particle-plus-measuring apparatus. If this is in a pure state and the wavefunction has the form (13.34), the expectation value of $\hat{Q}$ will be given by $\langle \hat{Q} \rangle$ where

$$\langle \hat{Q} \rangle = \int (A^* \alpha_z^* \chi_+^* + B^* \beta_z^* \chi_0^*) \hat{Q} (A\alpha_z \chi_+ + B\beta_z \chi_0)\, d\tau \tag{13.35}$$

and the 10^{20} or so variables required to describe the state of the particle and measuring apparatus are all assumed to be included in the volume element $d\tau$. Multiplying out (13.35) we get

$$\langle \hat{Q} \rangle = |A|^2 Q_{++} + |B|^2 Q_{00} + A^* B Q_{+0} + A B^* Q_{0+} \tag{13.36}$$

where

$$Q_{++} = \int \alpha_z^* \chi_+^* \hat{Q} \alpha_z \chi_+ \, d\tau \qquad \text{etc.}$$

In the case of the mixed state, $\langle \hat{Q} \rangle$ is the weighted mean of the expectation values corresponding to the states $\alpha_z \chi_+$ and $\beta_z \chi_0$. That is

$$\langle \hat{Q} \rangle = |A|^2 Q_{++} + |B|^2 Q_{00} \tag{13.37}$$

Comparison of (13.36) and (13.37) shows that these two results would be identical if $Q_{+0} = Q_{0+} = 0$; so, for this to be true when $\hat{Q}$ is any physical

operator, the mixed state and the pure state would be indistinguishable. Let us see how this might come about. One of the key features of any practical measurement is that there is an irreversible change in the measuring apparatus at some point. Thus when the counter clicks or the grain in the photographic emulsion blackens, an irreversible change occurs, accompanied by an increase in the entropy of the universe. By irreversible we mean that we cannot restore the initial state by any practical procedure. We note first that $|Q_{+0}|^2$ is proportional to the probability that a transition will take place between the states χ_+ and χ_0 under the influence of the operator $\hat{Q}$. But this would amount to returning the detector to the state it was in before the particle was detected, which would mean reversing an irreversible change! We might be able to do this by operating from outside the system, but this would increase the entropy of some other part of the universe, which would, in turn, become entangled with the system.

Decoherence has been closely studied in recent years. Some of this has been directed at the investigation of how quickly terms of the form Q_{+0} go to zero during a measurement. Models of measurement processes have been set up and in typical cases Q_{+0} decays to an incredibly small value in a very short time—typically around $10^{-10^{35}}$ after 10^{-6} seconds.

We can therefore conclude that, once decoherence has occurred, the statistical predictions for the properties of pure and mixed states are essentially identical. So has this solved the measurement problem? Yes, for all practical purposes, but we should have expected that. When Schrödinger posed the problem of his cat, he knew perfectly well that it would be completely impractical to recreate the initial state by some interference experiment. All decoherence really does is justify the practical theory of measurement that we started with, and most, if not all, points of principle remain unresolved. One of these is that, however improbable, there is always a theoretical possibility of a spontaneous reversal of an irreversible change. Perhaps more important is the fact that decoherence says nothing about collapse. The probabilities of obtaining various outcomes form just one feature of the measurement process. The other part of the postulate states that the wavefunction collapses into the eigenfunction corresponding to the result. This is no problem in the case of a mixture because the atoms are already in one or other of the eigenstates before the measurement is made: the measurement just has to tell us which particle is in which state. But in the general case discussed earlier, collapse means that the wavefunction must change in a way that is inconsistent with the time-dependent Schrödinger equation. To model collapse, not only must Q_{+0} and Q_{0+} vanish, but so must one or other of A and B at random. In practice, we assume this has happened after a measurement, but it is not included in the formalism. This is sometimes expressed by saying that collapse must be added 'by hand'.

Summarizing, there are two processes associated with measurement: decoherence and collapse and we still have not explained why collapse occurs. We will now discuss some of the ways in which this has been explored further. We define the *collapse time* as the time between the formation of a state such as

(13.34) and its collapse into one of its components, and we consider the relation between the collapse time and the decoherence time in various scenarios.

Many worlds

We first consider the possibility that the collapse time is infinite—i.e. collapse does not occur at all. At first sight, this seems to contradict everything we have just said about the importance of the collapse process. However, let us put aside this prejudice and consider what the universe would be like if there really were no collapse. Once decoherence has occurred and produced a state like (13.34), the whole system has been divided into 'branches', one of which corresponds to each possible measurement outcome. The key point is that not only can the initial state not be reconstructed, but there is also no way that the different branches can interfere with each other. Each branch evolves independently of the others, and as far as it is concerned, its future behaviour is just the same as if collapse had occurred and the other branches had disappeared. Another way of putting this is to ask how we could know that a particle is split into two delocalized halves as in the output of a SG apparatus. Any measurement we make on such a state simply entangles the apparatus with the particle to create a state of the form (13.34). The only way we know that the delocalized state ever existed is to perform an interference experiment, such as that illustrated in figure 13.7. Once this is impossible because of decoherence, the branches evolve quite independently and unaware of each other's existence.

These ideas, which were first put forward by H. Everett in the 1950s, involve a huge ontological cost. Not only does the apparatus branch, but so also does everything that interacts with it in the future—including human observers! As a result of looking at the counter I am split into two. One of me believes that the particle has been detected while the other thinks it has not, and my two selves can never communicate with each other to resolve their disagreement. Hence the term 'many worlds'. Every time an irreversible change occurs anywhere, the universe branches, so that there are now an unimaginably huge number of parallel 'worlds'. Many of these worlds will have appeared before human life evolved, but others will contain nearly identical copies of us.

Many-worlds theory addresses many of the problems associated with the quantum theory of measurement, but it does so in a particularly uneconomical manner. There is a fundamental postulate in science known as 'Occam's razor' which states that no theory should contain more postulates than are necessary to explain the observed facts. The idea of a near-infinite number of universes which can never interact with each other and whose existence can therefore never be verified seems to be an extreme breach of this principle. Nevertheless, this idea is taken seriously by some serious people and at least deserves further consideration of its merits.

First, many-worlds theory is a single theory. There is only the wavefunction governed by the Schrödinger equation and there is no need for collapse. Occam's

razor is only useful if we can agree on which assumptions we would most readily object to; some believe that many-worlds theory has an 'economy of postulates' that makes it preferable despite its 'extravagance with universes'.

Second, because the wavefunction does not collapse, the problems of non-locality discussed earlier do not arise. It is the act of *measuring* the state of one of the particles in an EPR experiment that causes the collapse of the wavefunction representing the other. If no collapse occurs, there is no change and no problem.

Third, it was a believer in many-worlds theory (David Deutsch) who was one of the first and main proponents of the possibility of quantum computing. As outlined in chapter 12, there are some calculations that would be carried out with enormous efficiency if a computer could be built that operated in reversible manner on the basis of unitary evolution. From a many-worlds point of view this gain arises from the possibility of carrying out a huge number of parallel calculations in different branches.

Fourth, a no-collapse theory is attractive to many of those interested in quantum cosmology. If we are to describe the whole universe by a wavefunction (as is done by those studying the very early stages of the big bang) it makes little sense to talk about an ensemble. The advantages of a single theory based on a wavefunction governed by the Schrödinger equation are particularly evident in this case.

However, many-worlds theory is subject to a major problem—quite apart from its ontological extravagance. This is the question of how we can talk about the probabilities of events when all possible outcomes actually occur. If a particle's spin is *either* up *or* down, it makes sense to attribute a probability to each outcome, and we can verify this by making measurements on a large number of systems and counting the fraction of the total in each. However, if the spin is simultaneously up *and* down and if there are separate copies of us associated with each possibility, what does probability mean? This question is particularly acute when we realize that the probabilities postulated in quantum mechanics are not related to the number of branches associated with each outcome. Thus, two branches evolve from a wavefunction such as (13.34) whatever the values of A and B (provided neither is zero), but the measurement postulate and experiment show that the probabilities are $|A|^2$ and $|B|^2$ respectively. A classical analogy is a flowing stream that splits into two parts of different size (figure 13.8). There may be more water going down one of the channels, but it is meaningless to say that the water is more likely to be in that channel, as it is actually in both. However, if we scatter floating objects in the stream, upstream from the division, we *can* say that they are more likely to be carried into one stream than the other. The equivalent of the floating object in quantum mechanics might be a Bohm particle (see earlier in this chapter), in which case probability is well defined. Indeed, the de Broglie–Bohm model can be thought of as a many-worlds theory, as both assume that the wavefunction follows the Schrödinger equation without collapsing. The rôle of the Bohm particle is then to pick out one of the branches as representing 'reality'.

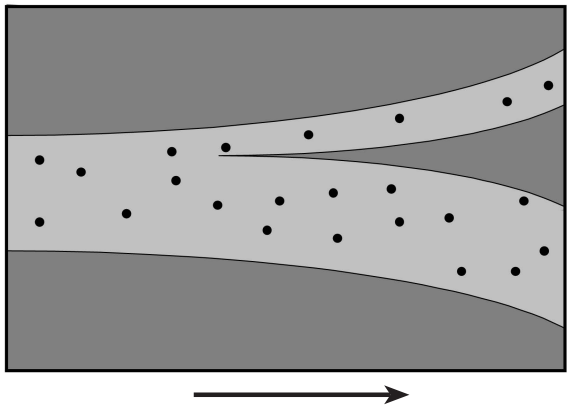


Figure 13.8. Water flowing along a stream divides into two channels. If small floating objects are scattered randomly onto the water above the splitting point, they will be more probably found in the wider channel. However, it is meaningless to say that the water is more probably in one channel or the other as it is actually in both. By analogy, probabilities can be straightforwardly assigned to particle properties in de Broglie–Bohm theory, but not to 'worlds' in many-worlds theory.

However, as we showed earlier, de Broglie–Bohm theory has serious problems of its own.

The problem of probabilities has been realized ever since many-worlds theory was invented and several workers in the field, beginning with Everett himself, have attempted to resolve it. However, no such proposal has yet commanded general acceptance, and the problem remains a substantial obstacle to believing in what is, anyway, a remarkably radical idea.

Spontaneous collapse

A completely different approach to the quantum measurement problem is to seek to modify the time-dependent Schrödinger equation in such a way that its solutions 'spontaneously' collapse into the expected states in appropriate circumstances. There would then be only one law governing time dependence and most of the problems would have disappeared. Thus objects such as counters and cats would be predicted to behave classically, while electrons or other microscopic particles would exhibit properties such as interference.

As we previously pointed out, the Schrödinger equation itself is a linear equation whose solutions describe the unitary evolution of the wavefunction, which forbids collapse. To include collapse, it must be modified by making it nonlinear—e.g. by including terms proportional to ψ^2 and/or higher powers of ψ. It can be shown that modifications of this type can be devised that will lead to collapse after some time, known as the 'collapse time'. Some of the earlier

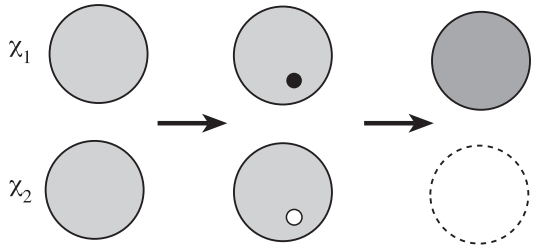


Figure 13.9. According to GRW theory, there is a finite probability of a particle collapsing into a localized state at any time. A macroscopic body is delocalized (picture on left) when one of the atoms collapses so that one of its possible positions is left empty, while the other is 'over-full' (central picture). To avoid the inevitable increase in energy, the whole object must collapse around the atom (picture on right).

suggestions for the form of such extra terms led to predictions about the behaviour of neutron beams that were tested experimentally and not observed. In general, a number of conditions must be fulfilled if any such theory is to predict collapse and be consistent with the known behaviour of quantum systems.

First, the collapse time must not be too short. Otherwise it would prevent interference between states that had been spatially separated for a longer time. Experiments performed on slow neutron beams have tested interference when the beams have been separated for the order of milliseconds with no sign of any spontaneous collapse.

Second, the collapse time must not be too long. As we have seen, when it is infinite, we essentially have a many-worlds theory. It follows that to avoid the problem of probabilities, collapse into one alternative *or* the other must happen sometime during the time taken to perform the measurement.

Third, collapse should occur in 'measurement-like' situations. That is, after the system has become coupled to some macroscopic apparatus, such as a pointer.

For some time it was believed that a spontaneous-collapse theory meeting these conditions was not possible. But a breakthrough in this area was achieved in the 1980s with the development of what has become known as 'GRW theory' after the initials of its three main protagonists.[4] This is based on the idea that the spontaneous collapse of a single particle contained in a macroscopic object necessitates the collapse of the whole. To understand this, imagine an object such as a pointer in a state where it is delocalized between two widely separated positions—i.e. a state such as (13.34) where χ_1 and χ_2 occupy different regions in space (cf. figure 13.9). Now suppose that the modified Schrödinger equation causes one of the atoms in the pointer to become spontaneously localized in a narrow region around some position in one region (say that occupied by χ_1); it therefore disappears from the corresponding region in the other. If nothing else

[4] G. C. Ghiradi, A. Rimini and T. Weber.

happened, the state would be a superposition of a complete pointer with an overweighted atom in one position along with one containing a missing atom in the other. However, the energy of this state would be increased by the energy required to create such hole. The only way that this can be avoided is if the whole pointer collapses along with the atom so that its wavefunction becomes χ_1, which is just what we are looking for.

A great strength of this idea is that it allows the collapse time for an isolated particle to be so long that it would be extremely unlikely ever to be observed, while ensuring that one of the many particles in the pointer would collapse in quite a short time. For example, if the collapse time for a single particle were to be around 10^5 years, the state of a pointer containing 10^{20} atoms would collapse in about 10^{-8} seconds. These principles have been developed into a mathematical theory and parameters can indeed be chosen that make it consistent with all presently performed experiments

There is a penalty to be paid for adopting a spontaneous-collapse model. This is that two new constants have to be postulated: the collapse time and a length specifying the size of the wavefunction after collapse. If these are fundamental constants, then the whole of physics is made more complicated and Occam's razor would encourage us to reject this unless it were supported by compelling experimental evidence.

To obtain positive evidence for spontaneous collapse, we would need to perform an experiment on a macroscopic body that would demonstrate interference if unitary evolution held, but not if collapse occurred. This means that we would have to delay the onset of decoherence by isolating the system from its environment. This is generally very difficult; for example, unless great care is taken to ensure that the two possible paths through an SG apparatus are kept constant to a high degree of accuracy (within about 10^{-6} m) the original spin function will not be reconstructed and the results of the experiment will be indistinguishable from one where the particles pass along one or other of the two possible paths without ever being delocalized. In fact the interference experiment illustrated in figure 13.5 has never been carried out with atoms, although the analogous experiment on polarized photons is relatively straightforward. Up to the end of the twentieth century, the largest objects shown to produce interference patterns after being passed through slits were Buckminster fullerene molecules—spherical molecules made up of 60 carbon atoms, which are many orders of magnitude too small for GRW collapse to apply. It was pointed out by A. J. Leggett in the late 1970s that, as the magnetic flux in superconducting quantum interference devices (SQUIDS) is associated with the collective coherent motion of a macroscopic number of electrons, an observation of quantum interference involving this quantity would show that quantum mechanics is applicable to a macroscopic object in this context at least. After two decades of experimental effort, nearly all quantum predictions for such systems have been checked and no evidence of collapse has been found. In late 2001, two samples of caesium atoms, each containing about 10^{12} atoms were maintained

in an entangled state for about 10^{-3} seconds before decoherence set in, and no collapse was detected. It is possible to choose values for the parameters of GRW theory so as to ensure that it is not falsified by any of these experiments, but most people will remain to be convinced unless and until collapse is observed in a situation where decoherence is forbidden.

Gravity

Apart perhaps from quantum mechanics itself, the greatest theory of modern physics is Einstein's general theory of relativity. This has successfully united a model of space and time with gravity to provide a consistent picture of the large-scale structure of the universe and its evolution from the big bang. The predictions of general relativity have been thoroughly tested and verified by a number of detailed astronomical measurements. However, major problems arise when we try to unite gravity with quantum mechanics. This is because the relativistic view of space–time and its curvature in gravitational fields is essentially incompatible with the linearity of quantum theory. It is generally believed that significant problems could only arise in situations where the gravitational fields are typically as strong as those believed to exist in the neighbourhood of a black hole—i.e. immensely stronger than anything experienced on earth. However, the cosmologist Roger Penrose has speculated that it may be the very weak residual effects of quantum gravity that cause the collapse of the wavefunction when the system is in a macroscopically delocalized state. This could conceivably provide a mechanism for GRW collapse, but will remain a speculation unless and until a full theory of quantum gravity is developed that contains this result, as well as others that can be tested experimentally.

The Copenhagen interpretation

We return now to what is generally accepted to be the conventional interpretation of the quantum theory of measurement. This was developed by Niels Bohr and co-workers in Copenhagen and is consequently referred to as the Copenhagen interpretation. There are several key ideas associated with this view.

First, the wavefunction has no counterpart 'in reality'. It is a purely mathematical object that enables us to make statistical predictions about the results of experiments.

Second, it is always incorrect to consider the quantum-mechanical system as separate from the measuring apparatus. If we call the set of eigenfunctions used to represent a state a *basis*, then any two orthogonal spin eigenfunctions—e.g. α_z and β_z or α_x and β_x—can be used as a basis for a given spin state. The Copenhagen interpretation states that the appropriate basis to use is determined by the details of the experimental arrangement. Thus a spin-half particle approaching an SGZ apparatus (including an appropriate counter) should be referred to α_z and β_z, while that consisting of a similar particle approaching an SGX must be

represented by α_x and β_x. If we remove the counter and bring the two paths together as in figure 13.6, we change the experimental set-up, and we should therefore not be surprised that the spin part of the wavefunction is α_x, rather than α_z or β_z. This aspect of the Copenhagen interpretation provides an alternative way of looking at the properties of entangled pairs of particles (see earlier in this chapter). Before a measurement, such a pair is a single quantum system whose properties cannot be accounted for in terms of the separate properties of the individual particles. Once we choose which spin directions to measure on each side, we define the appropriate basis for each particle. Because there is no reality in the absence of measurement, the arguments leading to Bell's inequality cannot be applied. These ideas were expressed by Neils Bohr when he said that the apparently non-local interactions were confined to 'an influence on the very conditions that define the possible types of prediction' that can be made.

Another important idea associated with the Copenhagen interpretation is that of *complementarity*. From this point of view some properties (e.g. position and momentum, x and z components of angular momentum, etc.) form complementary pairs, and it is an intrinsic property of nature that any attempt to define one of these variables precisely must lead to a complementary uncertainty in the other. For example, if a spin-half particle is in an eigenstate of S_x, to ask whether its spin is parallel or antiparallel to the z axis is clearly meaningless. In the same way, a proper understanding of quantum mechanics would imply that the question 'what is the position of a particle of known momentum?' is equally meaningless.

However, despite its great insights, the Copenhagen interpretation really does not address the problem of what constitutes a quantum measurement. Niels Bohr seems to have believed that the macroscopic apparatus is obviously classical with measurements corresponding to irreversible processes. Thus he does not appear to have distinguished between decoherence and collapse (neither word would have been used then) in contrast to Schrödinger, whose cat illustrates exactly this point. In the light of the work done since, a modern version of the Copenhagen interpretation might stress the irreversibility of the measurement process rather than the macroscopic nature of the apparatus. A basis should be chosen that is appropriate to the possible irreversible outcomes of the whole set-up. After the irreversible change has caused decoherence, the wavefunction should be taken as having collapsed into the form appropriate to the measurement outcome. As the wavefunction is a mathematical object with no direct physical significance, collapse is not a physical process. However, neither is it a mathematical process as we have to put it in 'by hand'. A spontaneous-collapse theory in which collapse and decoherence were coupled together so that the decoherence and collapse times were equal could account for this. Otherwise, the Copenhagen interpretation may mark the end of the 'Cartesian programme', by which is meant that Descartes' vision of a complete mathematical description of the physical universe may have reached its limit at the quantum measurement problem.

Subjectivism

The formal theory of quantum mechanics as outlined in chapter 4 is often thought to be an expression of the Copenhagen approach, but it was actually developed largely by von Neuman in the early 1930s. In contrast to Bohr, he did understand the measurement problem, but believed that collapse was a feature of human experience associated with 'psycho-physical parallelism'—by which is meant that our mind, or psyche, is not part of the physical universe, but is 'parallel' to it. This approach was further developed by E. P. Wigner. He pointed out that, in the last analysis, all we can know about the physical universe is the information that enters our mind through our senses and ends up in our consciousness. He therefore postulated that the collapse of the wavefunction occurs when the information enters a conscious human mind. Thus the particle, the counter and the cat are all in states described by expressions of the form (13.34), until a person opens the box and the information therein is transferred into a human mind; only at this point does collapse occur.

Although this theory can be made consistent with all the observed facts, it is generally considered unsatisfactory for a number of more or less obvious reasons. First, it relies heavily on the concept of the human mind or consciousness as something different in kind from the physical, material universe—including the brain. Certainly, many people, including some philosophers and scientists, believe this to be the case, but most would be unwilling to believe that the whole existence of the physical universe depends on this postulate: our aim should be to explain the natural world in objective terms consistent with it having an existence independent of our presence and interactions with it. Second, it pushes the whole problem into an inaccessible area; because if everything is 'in the mind' and this 'mind' is not a physical thing subject to investigation, then the whole of physics, and science in general, has no objective significance. Finally, it is difficult to explain the fact that different 'minds' generally come to the same conclusions about the results of physical measurements (in both classical and quantum mechanics) unless we allow the existence of an objective physical universe.

In recent years a combination of many worlds and subjectivism has been suggested. This proposes that there is no collapse; the wavefunction evolves according to the time-dependent Schrödinger equation and this is all there is in the physical universe. However, we conscious observers are incapable of seeing the world this way; it is in our nature that we can be aware of only one result of any measurement process so, although the others are still out there, we are unaware of them. It has been suggested that the problems of how different conscious observers always see the same results, points to all our individual consciousnesses being linked to some 'universal consciousness', one of whose jobs is to ensure that this agreement occurs!

13.5 The ontological problem

It should be clear by now that one of the fundamental problems thrown up by quantum mechanics in general, and the measurement problem, in particular, is the nature of reality—what is it that 'really exists' in the universe? What is our ontology?

The construction of a physical theory has been compared to the drawing of a map. In the same way that every point in a map has a direct counterpoint in the terrain being represented, so every parameter of a classical theory (particle position, momentum and energy, electromagnetic field etc.) has its counterpart in the real physical world. Quantum mechanics is quite different. The Schrödinger equation provides us with a map in terms of the wavefunction, but this has no counterpart in reality—at least from the orthodox viewpoint. Indeed, the only real events are the results of measurements and these are not part of the map, although their probabilities of occurrence can be derived from the map if we assume that it collapses from time to time.

Both hidden-variable theories and the many-worlds approach try to go beyond quantum mechanics and construct a map of reality. In the case of de Broglie–Bohm theory, both the wavefunction (as represented by the quantum potential) and the Bohm particle are assumed to exist and their real values can be predicted from the theory. If such a theory is ever developed with a self-consistent ontology that is in agreement with the principles of relativity, it will probably be greeted with enthusiasm and believed by most if not all. However, unless and until this happens, we will have to be satisfied with something less. In many-worlds theory, reality *is* the wavefunction and the map represents it completely. Essentially this is why surprisingly many people find it attractive, despite its ontological extravagance and its problems with defining probabilities.

A possible answer to the ontological problem is of course subjectivism: the only things that we know exist are our sense impressions, so perhaps we should define these as the only reality. Or we could adopt the approach known as positivism in which, because it is impossible to verify the existence or non-existence of anything beyond our sense impressions, we describe such questions as meaningless.

If, however, we stay with the orthodox approach to quantum mechanics, objective reality must be based on the idea of the irreversible measurement process. Perhaps it is only the irreversible changes in the universe which have a real, objective existence: any genuinely reversible changes are incapable of observation, and statements about their existence or otherwise are consequently meaningless. Moreover, by describing such reversible changes as 'incapable of observation' we do not necessarily refer to the intervention of any human observer: such reversible 'events' simply have no effect on the ensuing behaviour of the universe. Quantum mechanics can therefore be thought of as a theoretical system whereby we can predict, as far as this is possible, the sequence of irreversible events in the universe. In the process we use quantities such as

wavefunctions and talk about particles passing through slits without leaving any record, but it is only the irreversible changes which can be considered as having an objective existence. We have to live with the fact that the map does not represent reality and the Cartesian programme is not complete. Many still find this difficult to accept and the search for a better map will no doubt continue.

It was never the purpose of this chapter to provide answers to the conceptual problems of quantum mechanics, but to demonstrate that such problems do exist and that real questions can be asked in this area. In earlier chapters we have seen many examples of the success of quantum mechanics in predicting and explaining experimental results over a wide range of physics, and most students of the subject will probably continue to be content to use it as a theoretical tool in their study and research into particular phenomena. There is, after all, no dispute concerning the correctness of quantum mechanics in predicting quantities such as energy levels, transition probabilities, scattering cross sections, etc. Others will prefer to set the subject aside and do something else. In any case it is to be hoped that all students will understand that there are still some real problems in the grey area where physics and philosophy meet. The influence of natural science on philosophy has been very considerable, particularly in recent years, and the further consideration of the fundamental problems of quantum mechanics may very well have far-reaching effects on our understanding of our natural environment and, eventually, of ourselves.

Problems

13.1 Consider a wavepacket, approximated by a square pulse of width w, approaching a barrier in one dimension where the reflection probability is 2/3. According to de Broglie–Bohm theory, whereabouts in the packet must the particle be in order to be (i) reflected and (ii) transmitted?

13.2 If the z component and total spin of a particle are known, then the 'real spin' vector, if it exists, must presumably lie on a cone whose symmetry axis is parallel to z. Using this hidden-variable theory, show that in the case of a spin-half particle with positive z component, the probability of a component in a direction at an angle ϕ to the z axis also being positive is equal to

$$\begin{cases} 1 & \text{if } \phi < \cot^{-1}(2^{1/2}) \\ 1 - \pi^{-1}\cos^{-1}(2^{-1/2}\cot\phi) & \text{if } \cot^{-1}(2^{1/2}) \leqslant \phi < \cot^{-1}(-2^{1/2}) \\ 0 & \text{if } \phi > \cot^{-1}(-2^{1/2}) \end{cases}$$

Show that this result agrees with quantum mechanics if ϕ equals 0, $\pi/2$ or π, but not otherwise. If this theory were applied to measurements on correlated pairs, would it be a local or a non-local hidden-variable theory?

13.3 Consider a hidden-variable theory based on the model described in problem 13.2 with the additional condition that the result of a subsequent measurement of a component in the direction defined by ϕ is to be determined by the sign of the component of the 'real spin' in a direction at an angle ϕ' to the z axis, where in the case of a spin-half particle

$$\phi' = \cot^{-1}\{2^{1/2}\cos[2\pi\sin^2(\phi/2)]\}$$

Show that this theory produces identical results to quantum mechanics, but does not preserve locality when applied to measurements on correlated pairs.

13.4 A variant of the Schrödinger's cat experiment, in which the cat is replaced by a human observer,

was suggested by E. P. Wigner. This observer—known as 'Wigner's friend'—is not to be killed, however, but is to note and remember which of the possible quantum events occurs; subsequently the box is opened and Wigner makes a 'measurement' by asking his friend what happened. Discuss this procedure from the viewpoints of the different quantum theories of measurement, paying particular attention to the problem of wavefunction collapse.

Index